Design Space Constraints and the Cultural Taxonomy of European Final Palaeolithic Large Tanged Points: A Comparison of Typological, Landmark-Based and Whole-Outline Geometric Morphometric Approaches

David N. Matzig¹ · Shumon T. Hussain¹ · Felix Riede¹

Accepted: 18 May 2021 / Published online: 20 September 2021
© The Author(s) 2021

Abstract

The identification of material culture variability remains an important goal in archaeology, as such variability is commonly coupled with interpretations of cultural transmission and adaptation. While most archaeological cultures are defined on the basis of typology and research tradition, cultural evolutionary reasoning combined with computer-aided methods such as geometric morphometrics (GMM) can shed new light on the validity of many such entrenched groupings, especially in regard to European Upper Palaeolithic projectile points and their classification. Little methodological consistency, however, makes it difficult to compare the conclusions of such studies. Here, we present an effort towards a benchmarked, case-transferrable toolkit that comparatively explores relevant techniques centred on outline-based GMM. First, we re-analyse two previously conducted landmark-based analyses of stone artefacts using our whole-outline approach, demonstrating that outlines can offer an efficient and reliable alternative. We then show how a careful application of clustering algorithms to GMM outline data is able to successfully discriminate between distinctive tool shapes and suggest that such data can also be used to infer cultural evolutionary histories matching already observed typo-chronological patterns. Building on this baseline work, we apply the same methods to a dataset of large tanged points from the European Final Palaeolithic (ca. 15,000–11,000 cal BP). Exploratively comparing the structure of design space within and between the datasets analysed here, our results indicate that Final Palaeolithic tanged point shapes do not fall into meaningful regional or cultural evolutionary groupings but exhibit an internal outline variance comparable to spatiotemporally much closer confined artefact groups of post-Palaeolithic age. We discuss these contrasting results in relation to the architecture of lithic tool design spaces and technological differences in blank production and tool manufacture.

This manuscript is part of the topical collection *Cultural taxonomies in the Palaeolithic - old questions, novel perspectives*

Guest Editors Felix Riede and Shumon T. Hussain

Extended author information available on the last page of the article

Keywords Final Palaeolithic · Lithic technology · Geometric morphometrics · Outline analysis · Design space · Cultural evolution

Introduction

The robust identification of material culture variability remains an important goal in Palaeolithic archaeology, given that such variability is commonly coupled with interpretations of culture change, demography and adaptation. Material culture variability as seen in tool forms and assemblage structure is often readily agglomerated into higher-order units labelled ‘cultures’, ‘techno-complexes’, ‘groups’, ‘facies’, ‘industries’, ‘techno-units’ or the like (e.g. Gamble et al., 2005). For instance, in the Final Palaeolithic of Europe and the Paleoindian period of North America — what Williams and Madsen (2019) collectively refer to as the ‘Late Upper Palaeolithic’ — most such archaeological units are diagnosed, characterised and interpreted on the basis of artefact typology, chiefly that of lithic projectile points. Yet, critiques of such type-centred approaches abound (Bisson, 2000; Monnier & Missal, 2014), and recent applications of cultural evolutionary reasoning combined with computer-aided methods such as geometric morphometrics (GMM) have raised doubts as to the validity of many such entrenched groupings (Buchanan et al., 2020; Ivanovaitė et al., 2020; MacLeod, 2018; Serwatka & Riede, 2016). At the core of this emergent critique is the argument that traditional, essentialist typologies often fail to account for the internal variation seen in most qualitatively defined Palaeolithic artefact classes (cf. Shea, 2014; Wilkins, 2020) and that these typologies can easily reify preconceived differences (cf. O’Brien & Lyman, 2002). It has further been argued that the purported ‘cultural’ differences perpetuated through many traditional artefact typologies are more often than not the products of contingent disciplinary histories, rather than reflecting robust empirical realities (cf. Clark and Riel-Salvatore, 2006; Reynolds & Riede, 2019).

The analysis of artefact shapes through geometric morphometrics and cognate methods is by no means restricted to the Upper Palaeolithic, although the latter’s cultural taxonomic crisis may be particularly acute (see Reynolds & Riede, 2019). Acheulian handaxes (e.g. Costa, 2010; McNabb et al., 2018), the Middle Palaeolithic of Europe (e.g. Chacón et al., 2016; Serwatka, 2014) and elsewhere (Lycett, 2007), Holocene forager tools in South America (e.g. Charlin & González-José, 2012; Okumura & Araujo, 2014; Suárez & Cardillo, 2019) and later prehistoric metal axes (Wilczek et al., 2015), lances (Birch & Martín-Torres, 2019) and adornments (Vestergaard & Hoggard, 2019) have all been subjected to geometric morphometric assessments, often leading to a substantial critique of established typologies. There is little methodological consistency across these emerging case studies, however, and rarely have such studies re-analysed different datasets comparatively (see Monnier & McNulty, 2010, for an exception), making it difficult to evaluate their conclusions independently of their methodological choices. With the goal of better understanding the cultural evolution and adaptation of Final Palaeolithic (ca. 15,000–11,000 cal BP) hunter gatherers in Europe (Riede et al., 2020) and in an effort to move towards a reliable, case-transferrable analytical toolkit for doing so, we here present a multi-step comparative exploration of data capture, outline-based geometric morphometric

and clustering techniques. In the following, we re-analyse a series of published case studies using their lithic projectile point outline data in order to first develop and then validate our analytical protocol. This validation approach has four steps:

1. We benchmark whole-outline geometric morphometrics against a previously published landmark-based study aimed at discriminating two archaeological taxa in North America that were previously defined in terms of specific lithic point types (Buchanan et al., 2020).
2. We then similarly compare landmark-based geometric morphometrics and subsequent clustering as applied to a sample of Czech Bell Beaker projectile points (Petřík et al., 2018) with our whole-outline approach followed by hierarchical clustering. This comparison leads us to critically assess the choices of clustering algorithms.
3. Thereafter, we apply our protocol to a large sample of Late Neolithic/Early Bronze Age arrowheads from Northwestern Europe with well-constrained chronology (Nicolas, 2016). We show that our clustering protocol successfully replicates traditional techno-typological groupings and that the archaeological typo-chronology for this period, especially in North-West France, can be retrieved in a phenogram inferred from outline data using the neighbour joining (NJ) algorithm.
4. With this protocol in place, we finally turn to a dataset of Final Palaeolithic/earliest Mesolithic large tanged points from Europe (Ivanovaitė et al., 2020; Riede et al., 2019) in order to test groupings based on typological assessments, and to interrogate this sample in terms of the involved cultural evolutionary processes. We then compare the design space of the tanged points with the three previously re-examined lithic datasets by means of a combined principal component analysis (PCA). Doing so, we focus especially on the internal sum of variance which opens up new perspectives on rates of evolution and patterns of diversity in material culture in different time periods and cultural contexts.

The principal aim of our study is to develop a methodologically transparent protocol for the analysis of whole artefact outlines using GMM methods which can be used to establish and compare the shape variability of lithic artefacts through time and space. This approach facilitates the search for and the characterisation of structured variation within a given sample of artefacts therefore enabling the exploration of cultural transmission dynamics responsible for such variation.

Landmark and Outline Approaches in Geometric Morphometrics

Geometric morphometric approaches can be divided into methods that use landmarks and those that use trigonometric descriptions of whole outlines (Reyment, 1985; Temple, 1992), although the introduction of semi-landmarks and the dense placement of landmarks have led the two approaches to converge (MacLeod, 1999). Landmark approaches offer a straightforward way of delineating homologous structures, but their application also results in the loss of detailed shape

information. In addition, the a priori identification of homologous landmarks on artefacts can be difficult and inherently subjective unless unambiguous theoretical expectations are available (Lycett, 2009). Therefore, outline approaches offer an alternative, robust and information-rich way of systematically capturing artefact shape data. Accurate artefact outlines can also be extracted efficiently from widely available legacy data, especially from artefact line drawings that populate the archaeological expert literature and excavation reports.

In archaeology, geometric morphometric approaches in general constitute a major advance vis-à-vis earlier qualitative descriptions, typological assessment or linear measurements of artefact morphologies (e.g. Serwatka & Riede, 2016; Shott & Trail, 2010). The bulk of archaeological applications of GMM have so far relied on landmark-based approaches (Cardillo, 2010; Okumura & Araujo, 2018), although a surge of recent studies is demonstrating the utility of whole-outline approaches using so-called elliptical Fourier analysis (EFA; Kuhl and Giardina, 1982) and cognate approaches (Fox, 2015; Lipo et al., 2016; Wang & Marwick, 2020). Many of these studies also address shape change trajectories and discuss classification issues in Palaeolithic stone tools (e.g. Iovita, 2009; Leplongeon et al., 2020; MacLeod, 2018; Mesfin et al., 2020).

The application of phylogenetics to then interpret such classification in terms of cultural evolutionary dynamics is now well-established (Lipo et al., 2006; e.g. Lycett, 2010; Mace et al., 2005; Mace & Pagel, 1997; O'Brien et al., 2008). It remains contested, however, as to whether object outlines can capture phylogenetic signals and to what extent they convey social information. In evolutionary biology, Zelditch et al. (1995), for example, have claimed that while landmarks can serve as valid characters for phylogenetic reconstructions, outlines normally fail to do so. Still others suggest that especially in cases where the unambiguous placement of homologous landmarks is difficult, outlines can indeed record dynamics of evolutionary change (e.g. González-José et al., 2008; Polly, 2001; Smith & Hendricks, 2013). This being said, in Palaeolithic archaeology, shape is routinely mobilised as a significant trait for inferring cultural evolutionary processes such as change over time and group formation. Whole-outline morphometrics and landmark-based approaches have rarely been compared explicitly, however, so the relative performance of the two approaches remains poorly understood.

To address this key methodological issue, we here compare the results of prior landmark-based analyses of stone artefacts with the outcomes of geometric morphometric outline approaches. This benchmarking exercise begins with the re-analysis of Buchanan et al.'s (2020) study that used landmark data to resolve the question of a potential taxonomic division within Paleoindian projectile points. We then turn to the study of Petřík et al. (2018) who employed landmark-based GMM and hierarchical cluster analysis to investigate shape variation in Czech Bell Beaker points. Together, the methodological comparison of these two case studies facilitates the exploration of whole-outline GMM and its ability to address nested hierarchical classification. Building on these insights, we then apply our newly developed protocol to our third case, a substantial dataset of bifacially worked arrowheads from North-West France, Southern England and Denmark,

dating to between 4500 and 3700 cal BP (Nicolas, 2016). These arrowheads have previously only been examined from a traditional techno-typological perspective, but their regionality and chronology are well-established. This allows us to interpret our hierarchical classification methods in the light of cultural transmission dynamics. Equipped with this novel workflow, we finally turn to a sample of Final Palaeolithic large tanged points from central and eastern Europe in an effort to re-address their traditional classification and understand the cultural evolutionary dynamics within the design space framed by the larger assemblage of projectile points analysed here.

Outline Extraction and Transformation

For the extraction of outlines of multiple artefacts from either photographs or drawings, we developed an efficient workflow in the programming language *R* (R Core Team, 2020) based on the packages *EBImage* (Pau et al., 2010), *imager* (Barthelme et al., 2020) and *Momocs* (Bonhomme et al., 2014). If the quality standards of the original images are met and if there is no other information present on the image than the artefacts — such as artefact labels and scales — this workflow enables the fully automatic extraction of their outlines (Matzig, 2021). If heterogeneities exist in the original image (poor image quality, low contrast between artefact and background, shadow on photographs, etc.), they have to be subjected to minor manual preparation such as colour correction and manual thresholding/binarisation. The workflow is especially efficient for the extraction of individual shape information from images containing more than a single artefact, as it is the case for most archaeological catalogues and plate-rich publications. Each artefact in such a (binary) image is extracted using the *EBImage::bwlabel()* function and then automatically assigned a running number in addition to the identifier of the original table. Each extracted artefact is subsequently saved in JPEG while preserving the object's original size and position. If deemed relevant, this allows for retrieving the artefacts original measurements using a scaling factor from pixels to centimetres. With each individual artefact binarised, extracted, and archived, the images are ready for further processing in *Momocs*, where the function *Momocs::import_jpg1()* generates outline coordinates for each artefact. The workflow described here can easily be reproduced using the wrapper package *outlineR* (Matzig, 2021).

All outline datasets examined in this study were subjected to the same basic outline processing steps in *R*, also using the *Momocs* package. The prepared outlines were first centred and scaled. Then, EFA; was applied. In EFA, the shape contour is described in what is called harmonics — terms of sine and cosine curves of successive frequencies; we are here using the number of harmonics that describe 99.9% of the harmonic power. The data were then transferred into a PCA and used for further analyses, as described in detail below.

Protocol Development: Materials and Analytical Methods

This section introduces the three successive case studies from which outline data were extracted in order to replicate the original landmark-based or typological studies. We briefly present the context of each study, the main body of lithic data and our chosen approach, including the methods used to reproduce the original findings, to further explore the shape data and to test, and discuss and potentially expand upon previous findings and conclusions.

North American Late Paleoindian Bifacial Points

Using landmark-based GMM, Buchanan et al. (2020) addressed the relationship between two Late Paleoindian (ca. 12,700–11,400 cal BP) projectile point types from the Central United States: the so-called Goshen and Plainview variants (Fig. 1, 5–8). Due to an ambiguous chronology and overall similarity of these two point types, it continues to be discussed whether the two groups of artefacts should be kept separate or whether they should instead be consolidated into a single artefact group — a paradigmatic example of the splitter/lumper controversy (cf. Simpson, 1945). By applying landmark-based geometric morphometrics and statistical testing to a sample of 17 Goshen and 139 Plainview type points, the results of Buchanan et al. (2020) suggest that the two groups are difficult to separate and that Goshen should probably be incorporated into the Plainview group.

We tested whether the key results of Buchanan et al. (2020) can be reproduced using the outlines of the target artefacts, extracted from the original object photographs which were kindly provided by the authors of the original study. All outlines were processed and analysed as described above. The retrieved outline data, together with their assigned types, were then subjected to a permutational MANOVA, also known as PERMANOVA or nonparametric MANOVA, using the *vegan::adonis()*-function (Dixon, 2003). Pairwise distances within the function were calculated using Euclidean distance.

We further conducted a hold-one-out cross-validated linear discriminant analysis (LDA) using the *MASS::lda()*-function (Venables & Ripley, 2002) inside the *caret* environment (Kuhn et al., 2020) with the same data as for the PERMANOVA. The method's appropriateness for GMM analyses has been debated in the literature (Cooke & Terhune, 2015), but using the principal component (PC) scores for the calculation of the LDA heeds the assumption of a homogeneous covariance structure among groups. For the LDA, we first split the data into a training and test dataset with a ratio of 75:25%, respectively. Because Buchanan et al.'s (2020) sample sizes for the two groups are imbalanced, the training and test datasets were created using a stratified sampling approach, using the *splitstackshape* package (Mahto, 2019) in order to ensure that the type ratios were retained in the two subsamples. This training dataset was then used to initiate the cross-validated LDA model, while the test dataset was used to assess the performance of the said model. Furthermore, we subjected the complete dataset to a measure of relative shape distances. Proposed by Klingenberg and Monteiro (2005, 685), this technique provides 'a scalar measure

of the relative extent of shape differences, while taking into account that variation may not be isotropic'. Closely related to the one-sample version of what is referred to as Mahalanobis distance, this measure of asymmetry is designed to detect subtle shape differences that cannot be readily isolated in, for instance, PCA plots (cf. Klingenberg & Monteiro, 2005). Moreover, because this distance measure is scalar, it is amenable to univariate statistical analyses (Klingenberg & Monteiro, 2005). Therefore, we transformed our data according to the workflow described by Klingenberg and Monteiro (2005) to arrive at a measure of relative shape distances and assessed these via the Mann–Whitney U test.

Czech Bell Beaker Arrowheads

In order to assess whether our protocol is able to not only discriminate between two typologically pre-classified groups but also derive new, meaningful clusters of artefacts, we turned to the dataset of Czech Bell Beaker projectile points examined by Petřík et al. (2018; cf. Fig. 1, 11–14) who investigated the contribution of spatial distribution, raw material availability and curation intensity on shape and size variation among these lithic arrowheads. In a dataset of 194 projectile points from Moravia, Petřík et al. (2018) discriminated between several clusters of arrowhead shapes using landmark-based GMM and hierarchical clustering methods. Our aim, again, was to evaluate whether and to what extent our outline-based GMM protocol is capable of replicating those clusters.

The authors kindly provided us with the drawings of the respective stone artefacts, which were then processed following the workflow described above. Incomplete artefacts were removed from the sample, which reduced our outline dataset to 189 lithic artefacts. The first five PCs, describing 95% of the total shape variation, were retained. Petřík et al. (2018) similarly conducted a PCA on their landmark data and retained the first five PCs, which notably only explained 75% of the observed variability. Similar to Petřík et al. (2018), we then clustered the artefacts into groups based on a pairwise Euclidean distance matrix derived from the outline data's PC scores, employing the unweighted pair group method with arithmetic mean (UPGMA) algorithm (Sokal & Michener, 1958) in the base-R function *hclust()*. Following Petřík et al. (2018), we did not remove any outliers and cut the resulting tree in such a way as to also arrive at ten clusters. We then calculated the mean shape of the artefacts contained in each cluster for visualisation and direct comparison with the mean shapes retrieved in the original study.

The application of (phenetic) clustering methods, including the UPGMA algorithm, has a long history in archaeology (e.g. Clarke, 1968), not least in the Final Palaeolithic (e.g. Burdukiewicz, 1986; Schild et al., 1999). The number of clustering algorithms available is large, however, and explicit motivations or theoretical rationales for favouring one algorithm over another are rarely presented. Petřík et al.'s (2018) study is no exception here. Although hierarchical clustering as an unsupervised classification method generally suits the aims of this paper in so far as it represents a robust tool for an exploratory visualisation of group memberships

on different hierarchical levels, we follow Nakoinz (2013) in arguing that the choice of algorithm should always be motivated conceptually and/or theoretically. Our analytical aims align with ‘population thinking’ rather than essentialist typological preoccupations, rendering some clustering algorithms more appropriate than others (Table 1). With reference to now classic debates on essentialism and typological thinking on the one hand and materialism and population thinking on the other (Hull, 1965; Mayr, 1959; Sober, 1980), the pioneers of evolutionary archaeology have long argued against classification approaches that assume pre-existing mental templates or ideal forms (see, for instance, O’Brien & Lyman, 2002). More contemporary writing in the philosophy of biology is critically revisiting and correcting the original debate (Witteveen, 2015, 2018), however, and archaeological rhetoric should likely follow on towards a more nuanced epistemological understanding of variation and its generative processes (Tostevin, 2019).

Northwestern European Late Neolithic/Early Bronze Age Arrowheads

In order to extend our protocol to hierarchical relationships between objects and the clusters they form, we turned towards a sample of Northwestern European (i.e. North-West France, Southern England and Denmark) bifacially worked arrowheads published and examined by Nicolas (2016; cf. Fig. 1, 9–10), dating to the period of 4500–3700 cal BP. The original study contains detailed projectile point images, whose geographic provenance and chronological position are sufficiently constrained, and contains a techno-typological classification of these arrowheads projected on the period’s chronology, which is based on metric measurements and morphological observations. To investigate whether the coupled application of whole-outline GMM and tree-building methods can retrieve the cultural–historical sequence proposed by Nicolas (2016), we first extracted the outlines of 738 lithic arrowheads following our data capture protocol. Before we applied hierarchical clustering (Ward’s method) to the pairwise distance matrix (Euclidean distance) from all PC scores of the retrieved outline shapes, extreme outliers were removed (Online Resource Figs. 8 and 9). We determined the number of clusters k on the basis of the average silhouette value (Rousseeuw, 1987) calculated using the *NbClust* package (Charrad et al., 2014). The silhouette coefficient measures the average distance based on the observations of each point within a single own cluster compared to each point in the other clusters. The coefficient can thus be calculated for each observation in the whole dataset and averaged for each number of possible clusters. The highest average silhouette value indicates the optimal number of clusters k .

In addition to the hierarchical clustering approach based on Ward’s method, we also applied the NJ algorithm (Saitou & Nei, 1987; Studier & Keppler, 1988), another distance-based clustering method. This algorithm clusters, in each iterative step, the pair of taxa that lead to the largest decrease of the overall tree length. In contrast to the approaches discussed above, NJ approximates the minimum evolution method (Felsenstein, 2004) and thus does not produce ultrametric trees, meaning that different rates of (trait) evolution are allowed for each taxa. The branch lengths of the tree can then be interpreted as the taxa’s rate of evolution. Conversely,

Table 1 Selected phenetic clustering methods and their implications with regard to cultural evolutionary applications

Method	Implementation	Result	Implication	Cultural evolutionary inference
Single linkage (Florek, 1951; Sokal & Michener, 1958)	Clusters are formed based on the distance between the two closest members of each group	This leads to a chaining- or nesting-effect, connecting otherwise perfectly distinct groups to form a larger cluster	A single, real object represents a given cluster, grouping based on overall object similarity	Essentialism
Complete linkage (Sørensen, 1948)	Clusters will get linked based on their two most dissimilar members	Many small, artificially discretised clusters are created; is sensitive to outliers	A single, real existing object represents a cluster	Essentialism
Unweighted pair group method with arithmetic mean (UPGMA; Sokal & Michener, 1958)	Two clusters are merged based on the average of all possible distances between its respective group members	Clusters contain objects with similar variance	Compromise between single and complete linkage, mitigating their shortcomings (chaining effect, outlier sensitivity)	Interface between classical population thinking and typological thinking
Unweighted centroid clustering (UPGMC; Sokal & Michener, 1958)	An object centroid is computed based on the arithmetic mean of each object in the respective cluster. Those centroids are used to measure the distance between two clusters. After each step of agglomeration new centroids are calculated	Each cluster is represented by an idealised, abstract type	Idealised forms are created	Essentialism
Ward's Method (Ward, 1963)	Ward's method is based on the minimisation of square error. At each clustering step all possible pairs of clusters are compared and the one which results in the smallest variance is chosen	This algorithm has a tendency to impose compact clusters of equal size on a dataset	Emphasis is on variation	Population thinking

under the assumption of a constant rate of evolution in each taxa, the branch lengths correspond to the temporal distance from their evolutionary origin. However, due to the use of a pairwise distance matrix, this approach remains purely phenetic, as information about character state changes is unfortunately lost. We calculated the NJ trees using the *phangorn* package (Schliep, 2010) bootstrapped with 10,000 iterations on the same Euclidean distance matrix as used for the hierarchical clustering before, albeit restricting our analysis to the region of North-West France, which represents the largest and most densely sampled geographic area within this dataset.

Final Palaeolithic/Earliest Mesolithic Tanged Points

Finally, we turn to a previously collated dataset of Final Palaeolithic and earliest Mesolithic large tanged points from northern and eastern Europe. These data were assembled by Riede et al. (2019) and Ivanovaitė et al. (2020; cf. Fig. 1, 1–4), and their research historical context has been discussed in detail in these publications. Here, we have re-sampled this dataset in such a way as to retain only lithic artefacts of a defined named archaeological type (NAT) represented by more than three outlines. This resulted in the recovery of 253 complete large tanged point outlines from ten modern-day countries (Belarus, Belgium, Denmark, Germany, Lithuania, Poland, Russia, Sweden, the Netherlands and Ukraine) attributed to seven NATs traditionally used as cultural taxonomic diagnostics (Table 2, see Online Resource Table 1 and 2 for counts of NATs per chronozone and NATs per country, respectively).

The outlines of the dataset were processed and subsequently clustered in the same way as described in the case studies before, using the Euclidean distance matrix derived from all PC scores and Ward's method. The average silhouette value was used to determine the optimal number of clusters. As for the Late Neolithic/Early Bronze Age arrowheads, we furthermore inferred a phenogram using the NJ algorithm described above.

To assess the overall configuration of the design space occupied by Final Palaeolithic/earliest Mesolithic large tanged points, we then compared the retrieved point outlines with the shapes of the previously sampled North American Paleoindian, Northwestern European Late Neolithic/Early Bronze Age and Czech Bell Beaker lithic points. For the Late Neolithic/Early Bronze Age and Bell Beaker point data from Northwestern Europe and the Czech Republic, we drew a stratified subsample by region (using the *splitstackshape* package of Mahto, 2019) to yield, where possible, the same number of outlines as for the large tanged point dataset, and from thereon treated them as a single sample of Late Neolithic/Early Bronze points. We subjected this outline datasets to a combined PCA, again following the procedures described above and removing any identified outliers. Using all PC scores from this combined dataset, we calculated the shape disparity of each of the three samples. Disparity is defined here as the amount of total morphological variation, measured as the sum of variances within each of the dataset included in the combined PCA. For this, we used the *disparRity* package (Guillermé, 2018), bootstrapping the PCA data 1000 times. To test for statistical differences between the disparity values obtained for each of the three datasets, we conducted a pairwise Wilcoxon

Table 2 Outline counts per NAT that were retained for this study's analysis, their chronozone placement and approximate absolute chronology. NAT data derived from the Final Palaeolithic/earliest Mesolithic tanged point dataset published by Riede et al. (2019)

NAT	Modern-day countries	Chronozone placement	Absolute chronology (approx.)	N
Lyngby	Germany, Denmark, Belarus, Ukraine, Russia, Sweden, Poland, the Netherlands	Later part of Allerød/Bølling Complex	ca. 14–13 ka cal BP	137
Bromme	Germany, Lithuania, Denmark	Later part of Allerød/Bølling Complex	ca. 14–13 ka cal BP	58
Ahrensburgian	The Netherlands, Belarus, Denmark, Belgium, Germany	Younger Dryas/Preboreal transition?	ca. 13–12 ka cal BP	19
Altinovo	Belarus, Russia	Younger Dryas?	ca. 13–12 ka cal BP	16
Kaszaty	Poland, Lithuania	Younger Dryas?	ca. 13–12 ka cal BP	7
Krasnosillya	Belarus, Ukraine	Younger Dryas/Preboreal transition?	ca. 13–11 ka cal BP?	8
Grensk	Belarus	Younger Dryas/Preboreal?	ca. 13–11 ka cal BP?	8

Fig. 1 Examples of lithic artefacts subjected to 2D-GMM analysis. **1–4** Large tanged points from the European Final Palaeolithic; **5–8** North American Paleoindian points; **9–14** Late Neolithic/Early Bronze Age arrowheads from central and Northwestern Europe. **1** Large tanged point with minimal distal edge modification from Eskebjerg, Denmark (Fugl Petersen, 2001, Fig. 12, 1); **2** large tanged point with direct lateral retouch from Eskebjerg, Denmark (Fugl Petersen, 2001, Fig. 12, 5); **3** triangular tanged point with one-sided concave basal modification from Golßen, Germany (Winkler, 2010, Abb. 16: 18); **4** tanged point with minimal distal edge correction from Golßen, Germany (Winkler, 2010, Abb. 16: 15 after Geupel, 1987); **5** bifacial Plainview point with concave base from Ted Williamson, US (courtesy of B. Buchanan); **6** bifacial Goshen point with concave base from Hell Gap, US (courtesy of B. Buchanan); **7** bifacial Plainview point with straight base from Milnesand, US (courtesy of B. Buchanan); **8** bifacial Plainview point with straight base from Milnesand, US (courtesy of B. Buchanan); **9** bifacial triangular, winged-stemmed arrowhead from Tumulus de Cruguel, Morbihan, France (Nicolas, 2016, Planche 73: FR-58-08); **10** bifacial arrowhead with concave base from Blære, Års, Denmark (Nicolas, 2016, Planche 170: DK-05-07); **11** bifacial arrowhead with concave base from Jezefany-Maršovice, Czech Republic (Olivík, 2009, Tab. 38: 6) after Kopacz et al., 2009; **12** bifacial arrowhead with concave base from Jiřikovice, Czech Republic (Olivík, 2009 Tab. 38: 7) after Kopacz et al., 2009; **13** arrowhead with heavy, circumferential edge retouch from Ledce II, Czech Republic (Olivík, 2009, Tab. 38: 8) after Kopacz et al., 2009; **14** arrowhead with heavy, circumferential edge retouch from Ledce II, Czech Republic (Olivík, 2009, Tab. 38: 9) after Kopacz et al., 2009). No scale

signed-rank test with Bonferroni correction to decrease the likelihood of Type I errors for multiple testing.

Protocol Evaluation: Results

In this section, we present the results of our whole-outline 2D analysis of three selected lithic point samples (Paleoindian Goshen and Plainview points, Czech Bell Beaker and north-western Late Neolithic/Early Bronze Age arrowheads) and compare them with the conclusions of their original studies in order to assess the performance of our approach in reproducing the results of landmark-based approaches or traditional typological reasoning, respectively. While we recognise that different approaches to describing artefact form, shape, and geometry necessarily come with different sets of advantages and disadvantages and hence complement each other in varying ways, the purpose here is to evaluate whether and to what extent outline-based geometric morphometric methods offer a promising, feasible, effective and efficient way to capture the variability of artefact shape of particular stone artefact classes. Acknowledging the discussion about evolutionary information contained in outline data briefly mentioned in the beginning (i.e. Zelditch et al., 1995, 2001; MacLeod, 1999, 2001), we further test to what extent this approach is suited to infer cultural evolutionary relationships using two-dimensional outline data.

Reproducing the Results of Buchanan et al. (2020)

In our whole-outline GMM analysis of Buchanan et al. (2020), the first two PCs can be interpreted to describe the artefacts' width (PC1 = 61.3%) and symmetry (PC2 = 24.8%), respectively (Fig. 2). Except for a few outliers, the two types show considerable overlap. While parametric tests such as the Mahalanobis distance between

the group centroids (using the first six PCs, $D^2 = 1.06$) or Hotelling's generalised T^2 -statistic ($p = 0.021$, $T^2 = 0.104$) would suggest statistically significant differences, they assume multivariate normal distribution patterns which is not supported

by the data. In contrast, a nonparametric PERMANOVA test with 10,000 permutations resulted in $p = 0.338$ ($F = 1.071$). Thus, H_0 (the means of all groups are equal) could not be rejected, indicating that the groups are statistically indistinguishable. Likewise, the distributions derived from the measure of relative shape distances (Klingenberg & Monteiro, 2005) subjected to a Mann–Whitney U test yielded a non-significant result ($p = 0.46$; $W = 1312$), again suggesting that group means cannot be separated. For the results of the LDA, all Goshen-type points of the test dataset were classified as Plainview type. Here, all Plainview type specimens from the test dataset were classified correctly. In sum, our analysis confirms the results of (Buchanan et al., 2020) and simultaneously demonstrates that outline-based approaches are a robust alternative for capturing lithic artefact shape vis-à-vis landmark-based approaches. Having said this, it should be noted that the Goshen point sample is fairly limited, comprising only 17 specimens and therefore places the discriminative power of the statistical analyses on relatively weak grounds.

Reproducing the Results of Petřík et al. (2018)

As shown in the original study by Petřík et al. (2018), the first two PCs of our replication study cover variability of point width and base shape, respectively. Yet, while PC1 in the original study only explained ~34% of the total shape variation, it explains 76% in our outline-based analysis. Petřík et al.'s PC3 reflects object symmetry and explains 10% of variation. In our study, symmetry is instead represented by PC5 and explains merely 3% of the total variation. Both approaches thus identify the same general dimensions of shape variation, but the whole-outline method seems to accord greater importance to lithic object width and the relative convexity of the artefacts' lateral edges.

Inspection of the PCA plot of the outline data suggests that PC1 explains such a large amount of variation due to four outliers which belong to Clusters 6 and 8 (Fig. 3). We suggest that those outliers are responsible for Petřík et al.'s (2018) scree plot, indicating a much lower optimal number of clusters (see Petřík et al., 2018, supplementary Fig. 3). Our cluster analysis using the UGPM algorithm thus indicates that we are largely dealing with a single point type, characterised

Fig. 2 Scatterplot of the first two PC scores of Goshen and Plainview point outline data with their respective 95% confidence interval ellipses created with *ggplot2* (Wickham, 2016)

Fig. 3 The ten clusters generated from the whole-outline GMM re-analysis of Bell Beaker arrowheads from Petřík et al. (2018). **A** The generated mean shapes and the associated UPGMA dendrogram. **B** Scatterplot of the first two PC axes. **C** Individual clusters and their respective artefact shapes

by a concave base and the absence of a tang and with variation in overall width and base concavity (see especially Online Resource Fig. 7). Wider points with a flatter base are grouped in Cluster 1, whereas the narrower points with more pronounced concave bases fall into Cluster 4. Clusters 2 and 3 represent stemmed points, whereas clusters 6, 8 and 9 seem to gather outliers, as do Clusters 7 and 10, which represent asymmetric points with a single wing on their right side. We followed Petřík et al. (2018) in dividing the dataset into ten clusters. In general, the mean shapes of the retrieved ten groups are comparable to those derived in the original study. While we cannot further quantify the degree of match between the results of the landmark-based and whole-outline GMM methods, they produce similar overarching trends: two main clusters surrounded by many outliers (see Petřík et al., 2018, Table 1). It is worth noting here, however, that the outline-based approach appears to be more sensitive to outliers, potentially leading to heavy distortions in the PCA.

Fig. 4 **A** Pruned dendrogram and derived mean shapes of the respective clusters based on Ward's method using the outlines of the stone arrowheads published by Nicolas (2016). **B** PCA scatterplot of the first two PCs, coloured according to cluster attribution. **C** Overview of outlines contained in each cluster

Reproducing the Results of Nicolas (2016)

The average silhouette value ($k = 9$; Online Resource Fig. 12) derived from the point data of Nicolas (2016) results in the dendrogram shown in Fig. 4 which is based on Ward's method. The shapes contained within the retrieved clusters closely mirror Nicolas' (2016, Fig. 29) original typological divisions (Fig. 4; Online Resource Fig. 11). The types proposed by Nicolas (2016, Fig. 29) align along two axes, the first representing the shape of the arrowheads' base (ranging from hollow to concave to barbed); and the second conveying the presence and form of the barbs (ranging from pointed to so-called slanted ones). Regarding the shape variation captured by the analysed outlines, PC1 (54%) represents a combination of axis one and axis two of the author's original type-based classification: it mainly describes the arrowheads' base and ranges from a highly concave base with rounded, downwards pointing barbs and straight lateral edges to arrowheads with a very long stem and long, slanted barbs with generally more curved lateral edges. PC2 (24%) yields information about the barb-length-to-tang-length ratio and the global width of the

artefacts, whereas PC3 (9%) describes the overall barb-length-to-tang-length ratio, with artefact width only playing a minor role. PC4 (4%) captures the configuration of the base with an emphasis on the overall length of tang and barbs. The symmetry along the main tool axis, especially concerning the unevenness in barb length and the extent of the hafting indentation, is represented in PC5 (2.5%).

Overall, the artefacts are best described as belonging to only two clusters — those with tanged versus those with concave bases — with a further subdivision of the tanged points into plump and less symmetrical arrowheads with short tang and barbs (Clusters 3 and 4) and more slender artefacts characterised additionally by their elaborated proximal part and high overall symmetry (Clusters 5, 6, 7, 8, 9). Importantly, this pattern aligns with the geographic distribution of the artefacts within the clusters (Fig. 5). While Clusters 1 and 2 almost exclusively appear in Denmark, Clusters 5, 6, 7 and 9 occur mainly in France. Clusters 3, 4 and 8, in contrast, are found both in France and Britain, whereas the few British arrowheads in Cluster 8 form a distinct subsample within this cluster.

In addition to these well-defined geographic patterns, the phenogram inferred by the NJ algorithm (Fig. 6) structurally replicates the typo-chronology established in the original study. Assuming a constant rate of trait evolution (see above), the branch length of the X-axis reflects the temporal distance to the taxon's ancestors. The tip labels are coloured according to archaeological sites and named according to their catalogue entries; the coloured column on the right represents the time period to which the corresponding arrowheads have been dated. Following this, the clade highlighted in yellow captures the transition from Bronze A1 Stage 1 over Bronze A2 Stage 2 to Bronze A2 Stage 3 (cf. Nicolas, 2017, Fig. 10). The clade coloured in light green captures the transition from Bronze Age A1 Stage 1 (close to the root of the tree) to Bronze A2 Stage 2 (with the longest branches). The clade between the clades highlighted in green and yellow exclusively contains artefacts from Bronze A1 Stage 1. The lower most clade containing mostly Bell Beaker arrowheads yields a more ambiguous pattern.

Protocol Application: Final Palaeolithic/Earliest Mesolithic Tanged Points and the Evolution of Lithic Design Spaces

Having demonstrated our outline-based protocol's ability of capturing the main features of artefact shape variability in ways comparable to previous typo-technological or landmark-based approaches, we now apply the protocol to the sample of large tanged points from the Final Palaeolithic and earliest Mesolithic of northern and eastern Europe assembled by Riede et al. (2019) and Ivanovaitė et al. (2020). The purpose of revisiting these data is to apply the 2D outline-based approach developed here to explore the design space and potential cultural evolutionary signatures of this artefactual *Leitfossil* in relation to the proposed cultural taxonomic schemes and to compare the resulting patterns with the other three datasets analysed here.

The shape space of tanged points in the dataset is mainly described by two first PCs. A total of 50% of the shape variance is described by PC1, reflecting symmetry along the main tool axis and the position of maximum width (Fig. 7). PC2

Fig. 5 Geographic distribution of the derived Late Neolithic/Early Bronze Age arrowhead clusters based on Ward's method, excluding outliers. Colour coding is identical to Fig. 4

captures 27% of the shape variance and charts the relationship between the tang area and the distal part of the points as well as their slenderness. PC3 describes 6% of the recorded variance and tracks different distal edge configurations (straightness, convexity and convergence). Only 4% of the shape variance of tanged points is explained in terms of the relative length of the tang and the transversal curvature of the distal part. The overall shape space of these tanged points is relatively coherent with a dense centre and no extreme outliers (see Fig. 7A, C, E).

The cluster analysis of tanged point outlines using Ward's method yields little evidence in favour of unambiguous and well-defined groupings. The comparatively short branch lengths of the dendrogram indicate strong overall similarity in the outline dataset and limited evidence for higher-order groupings (Fig. 8). Based on the

Fig. 6 Phenogram based on the NJ algorithm using only the outlines of the artefacts from North-West France published by Nicolas (2016). The tip labels are coloured according to their respective sites and named after their catalogue entry; the coloured column on the right represents the time period into which they have been dated. Highlighted in yellow and green boxes are the clades that mirror the original study's typo-chronology particularly well

Fig. 7 Shape space of the Final Palaeolithic/earliest Mesolithic tanged point dataset. **A, C, E** The scatterplots of PC axis 2, 3 and 4, respectively, combined with PC 1, colour-coded in accordance with the clusters in Fig. 8. The associated shape variation with each PC is depicted in **D**. **B** The proportion of variance explained by the respective PCs

derived silhouette plot (Fig. 8B), the local optimum of clusters is $k = 10$ and the artefact composition of each cluster is provided in Fig. 8C. The resulting lower-level groupings of tanged lithic points (Clusters 1–10) further exhibit pronounced internal shape variability. Intra-group shape variance is lower, however, than overall inter-group shape variance, suggesting that these groups reflect general shape tendencies within a single population or design space. This observation is further supported by the unstructured geographical distribution of the artefacts (Fig. 9). The only exception here are perhaps the distinctly shaped artefacts of Cluster 9 from Belarus and Russia.

Paralleling the application of the NJ algorithm to the French Late Neolithic/Early Bronze Age points (cf. Fig. 6), we seek cultural evolutionary signals in the Final Palaeolithic/earliest Mesolithic large tanged point sample. Based on the methodology

Fig. 8 **A** Pruned dendrogram (Ward's method) based on the Final Palaeolithic/earliest Mesolithic sample of tanged points. The mean shapes of the clusters are placed at the tree's tips, the individual artefacts contained in each cluster are depicted in **C**. **B** The average silhouette value for the dataset if partitioned into a different number of clusters

developed here and the results obtained on our training samples, we expected that if a coherent cultural evolutionary signal is inherent in the shapes of the large tanged points, then these should cluster chronologically, regionally or both. Yet, the inferred large tanged point phenogram yields no obvious association between chronological position (relative chronozone) and their branch lengths, or by their geographic origin or prior cultural taxonomic assignment (Figs. 9 and 10), as was the case for the French Late Neolithic/Early Bronze Age arrowheads.

In a final step and in an attempt to assess the significance and configuration of the global shape space constraints of our Final Palaeolithic/earliest Mesolithic tanged points, we compared their outlines with the outlines of the North American Late Paleoindian and the combined Northwestern European Late Neolithic/Early Bronze Age and Czech Bell Beaker point and arrowhead samples examined above in a single, combined PCA. The PCA analysis produces two distinct and well-separated groups (Fig. 11). Almost 90% of the total shape variance is explained by the first PC, however, most likely reflecting different overall design solutions. The extremes of PC1 are occupied by forms which can be described

Fig. 9 Geographic distribution of the derived Late Palaeolithic/earliest Mesolithic tanged points dataset clusters based on Ward's method, excluding outliers. Colour coding is identical to Fig. 8

as oval at the one end and forms trending towards more triangular shapes on the other. Another key dimension of the shape space is the configuration of the base and the position of the maximal width which is represented in PC2 (6%). Merely describing 2% of the recorded shape variance, PC3 reflects the degree of symmetry, although do recall the substantial difference here between the highly symmetric Paleoindian and Late Neolithic/Early Bronze Age points and the strongly asymmetric Final Palaeolithic ones. PC4 (2%) captures slenderness and edge curvature. PC1 discriminates between the large tanged point sample at one extreme and the Paleoindian and Late Neolithic/Early Bronze Age points from on the other extreme. The Paleoindian points exhibit the greatest internal homogeneity across all PC axes suggestive of either strong functional constraints, strongly normative cultural transmission processes, or possibly a combination of these. The variance of Late Neolithic/Early Bronze Age arrowheads is greater and primarily explained by PC2. It is important to recall, however, that the variation within this sample was highly structured with nested clusters (Fig. 4). The Final Palaeolithic/earliest Mesolithic large tanged points, in contrast, show a PC variance comparable to the combined Late Neolithic/Early Bronze Age sample in the absence of geographical structure. These diverging trends are also reflected in the overall sum of variance (disparity) within each of the three samples (Late Paleoindian, Late Neolithic/Early Bronze Age and Final Palaeolithic/earliest Mesolithic large tanged points), yielding statistically significant results regarding their pairwise

difference (pairwise Wilcoxon signed-rank test $p < 0.001$, Fig. 11F). The high disparity in the Late Neolithic/Early Bronze Age dataset is explained by its strong internal and also regional structure. When split into regions, the region's internal sum of variance becomes even lower and approaches the variance value calculated for the tanged point dataset (Online Resource Fig. 13F).

It is also worth noting that the retrieved groupings in Fig. 11 are distinguished by the mode of tool production and retouch: the large tanged points are unifacial and often represent only locally retouched and edge-modified pieces, whereas both Paleoindian and Late Neolithic/Early Bronze Age points are bifacially modified. The position and spread of the Late Neolithic/Early Bronze Age arrowheads within the larger design space clearly shows that these forms are underpinned by the regulative idea of triangular and cordiform shapes and a marked trend from concave-winged to stemmed-winged proximal designs. The design of the proximal part also appears to be the distinguishing feature of the examined Paleoindian points, yet with an emphasis on relatively straight and horizontally undifferentiated basal configurations. The shape of the tanged points, by contrast, is defined by a more or less developed and elongated tang.

Discussion and conclusion

This study aimed at evaluating whether and to what extent outline-based geometric morphometric methods (GMM) are suited to capture the variability of artefact shape within stone artefacts from different chronological and cultural contexts in an efficient and robust way and whether and to what extent outline data are suited to reproduce previously established typo-chronological patterns. We have first benchmarked our 2D outline approach to lithic artefact shape analysis against three previously published datasets, two of which used landmark-based GMM. When compared to the landmark-based results of the Paleoindian (Buchanan et al., 2020) and Bell Beaker projectile point (Petřík et al., 2018) studies, our results demonstrate that an outline-based approach can reliably reproduce the primary dimensions of shape variability and the general structure of projectile point design space. We therefore conclude that the main advantage of outline-based approaches over landmark-based analysis, especially when integrated into semi-automated digital extraction protocols as developed here (Matzig, 2021), is the possibility of: (a) rapid data capture from published drawings or photographs; and (b) of circumventing the issue of defining homologous landmarks.

The case of Late Neolithic/Early Bronze Age arrowheads from Nicolas (2016) demonstrates that appropriate algorithm choice enables the retrieval of meaningful hierarchical clusters from artefact outlines which mirror previously and independently established archaeological types and so help to inform the interpretation of the geographic distribution of these artefacts. Moreover, and on the basis

Fig. 10 Inferred NJ phenogram of tanged points from the Final Palaeolithic and earliest Mesolithic of northern and eastern Europe. The tip label colour indicates the time period into which the artefacts have been dated, while the column on the right represents the NAT

of a dated subsample of the same dataset, we were able to infer a phenogram that closely resembles the original typo-chronology elaborated by Nicolas (2016).

Based on these results, we were able to confidently turn towards the Final Palaeolithic/earliest Mesolithic large tanged points dataset, which has been used in several studies so far (cf. Riede et al., 2019; Ivanovaitė et al., 2020) which were concerned with the questioning of the validity of so-called NATs characteristic for this time period. Using the clustering approach tested in the Late Neolithic/Early Bronze Age case study before, the retrieved clusters of tanged points do neither converge with the NATs to which they are traditionally ascribed, nor does the geographical distribution of the clusters point towards any patterns of regionalisation. While keeping the skewed distribution of artefact numbers per NAT in this dataset in mind — as well as the compelling results obtained from the other datasets analysed above — we conclude that the lithic artefacts of the different NATs form a comparatively homogeneous group and that the documented variability likely reflects residual variability related to blank production and original blank morphologies. At least a part of the documented outline variability of the large tanged points, in other words, is probably unrelated to the tool production itself.

The disparity analysis puts the variability of the large tanged points dataset into further perspective. Despite covering the largest period of time and broadest total geographical distribution, the tanged point dataset from the Final Palaeolithic/earliest Mesolithic exhibits an internal sum of variance comparable to each of the separate Late Neolithic/Early Bronze Age arrowhead subsamples from the UK, France and Denmark which document pronounced spatiotemporal differences despite geographic adjacency. Furthermore, the inferred phenogram of the large tanged point dataset does not indicate any chronological pattern (judged by either their relative chronozone attribution or their position in their imputed archaeological type names which are, respectively, rather broad and fuzzy and have previously been contested; cf. Reynolds and Riede, 2019). In relation to confounding factors such as the low chronological resolution of the sample, imbalanced artefact selection and low number of analysed specimen as well as a limited number of NATs, these findings should be seen as a starting point for further analysis.

The results of our analysis demonstrate that quantitative outline approaches to artefact shape spaces can, in principle, help to inform our understanding of cultural evolutionary dynamics of lithic forms, as is, for example, demonstrated by our replication of Nicolas' (2016) arrowhead-type sequence. Yet, we also caution that while overall shape may reflect the broader design space configurations of prehistoric projectile points including large tanged points and arrowheads, technological processes and attributes may yet be as informative or more so in terms of cultural transmission dynamics (e.g. Tostevin, 2019). Shape is only one dimension of point design, and understanding the morphological variability of pointed tools ultimately requires

Fig. 11 Combined shape space of all of the datasets studied. **A**, **C**, **E** The scatterplots of PC axis 2, 3 and 4, respectively, combined with PC 1. **D** The shape variation associated with each PC. **B** The proportion of variance explained by the respective PCs. **F** Boxplots visualising the sum of variance (disparity) for each of the tool groups studied. In each panel, the colours represent the three populations under study, as labelled in **F**

paying close attention to the interaction of shape-related variables with other key aspects of tool-making and tool design (cf. esp. Gowlett, 2006; Richter, 2013), such

as object biography and allometry, blank production, manufacturing techniques, and the strategies and goals of tool modification and curation. In tool designs largely centred on unifacial blanks, the relationship between blank production and tool manufacture may partially pre-determine point variability (cf. Migal, 2007). There is a clear difference in the amount of immediate shape control afforded by tools manufactured on *débitage* blanks and those made on *façonnage* products, especially when secondary modification targets only a part of the blank matrix and therefore only makes a comparatively small contribution to the whole outline of the final tool form. In contrast to bifacial shaping which generally enables high levels of overall shape control, it must thus be expected that at least some residual shape variability of tools made on these blanks reflect the variability of forms supported by the underlying blank production technology. In other words, a large portion of the tool outline is not directly controlled by the modification stage of the operational sequence but instead indexes the relationship between lithic production technologies and blank selection. This recognition has significant implications for how to interpret differences in the degree of shape normalisation detected by outline-based approaches.

In comparison to the large tanged points, the post-Palaeolithic arrowheads from central and Northwestern Europe analysed here attest to a tool design converging on broadly cordiform and triangular shapes. The degree of base differentiation (e.g. winged or winged and stemmed) appears to be the main design feature. The North American Late Paleoindian points converge on broadly the same design space yet trend towards straight, non-differentiated proximal parts, thus achieving an integration of ‘prehensile’, ‘transmissive’, and ‘transformative’ tool parts (sensu Lepot, 1992; Albrecht & Müller-Beck, 1988; Boëda, 2013; Richter, 2013). This regulative idea, in turn, may be responsible for the comparatively low shape disparity recorded in this point sample, indicating well-established conventions of global tool morphology and possibly also elevated levels of tool production and modification normativity. The Final Palaeolithic/earliest Mesolithic tanged points from northern and eastern Europe populate the other half of the total design space as their oval form tendency with pronounced tangs is clearly different from the mainly triangular shape tendencies of the North American Paleoindian points and the European post-Palaeolithic arrowheads. This general segregation of the design space also becomes evident in the PCA scatterplot (cf. Fig. 11), suggesting that proximal structural differentiation is another discriminating factor together with the degree of symmetry and point lateralisation. Unsurprisingly, the design of the studied tanged points is clearly centred on a more or less pronounced and well-defined ‘receptive’ or ‘prehensile’ tool part — the tang — leaving the distal ‘transmissive’ and ‘transformative’ tool parts to vary strongly.

In thinking about the generative processes behind the differences in the design between Final Palaeolithic/earliest Mesolithic tanged points and later prehistoric arrowheads, it is important to not only consider different design imperatives but their larger archaeological contexts as well. The Czech Bell Beaker and Northwestern European Late Neolithic/Early Bronze Age arrowheads, for example, make their appearance within a socio-political context in which knapped stone technologies co-exist with pottery technologies, polished and ground stone technologies and early metal production (e.g. Sarauw, 2008). As a consequence, design imperatives and

object norms are potentially at least in part also shaped the interplay with these other technologies, providing space for functional diversification and social status negotiations within a socio-political economy of object use and significance, including varying degrees of symbolic investment. Given that primary subsistence-related and economic roles may then be overtaken by other key technologies which potentially even outperform lithic tools in these respective functions, lithic tools can acquire potent sociocultural, ideological, and cosmological value or are relegated to specialised domains of human behaviour such as warfare, social stratification or territoriality (e.g. Dietrich et al., 2019; Gopher, 1994; Whittaker, 1994). Such intra-society dynamics create the cultural and technological space for a marked diversification of artefact morphotypes. Lifting the morpho-functional but also the labour constraints of lithic points may incentivise the exploration of the available and adjacent design space even further and therefore accelerate the evolution of artefact forms.

This situation is contrasted by the Final Palaeolithic/earliest Mesolithic points. The morphotype of the large tanged point (≥ 15 mm and >5 g), which in Europe continues to be treated by some scholars as a classic culture-historical *Leitfossil* of the Final Palaeolithic (ca. 15,000–11,000 cal BP), is found in the archaeological record of many regions and periods (e.g. Pesesse & Flas, 2012; Seong, 2008), suggesting a global design optimum on which lithic point design converges multiple times even in geographically unrelated areas (cf. O'Brien et al., 2018). One explanation for their lack of chronological and regional specificity may thus be sought in the role of these large tanged points as functional components of larger systems of hunting weaponry and their associated subsistence and mobility strategies. For example, the primary occurrence of such massive lithic points in higher latitudes may relate to their involvement in specialised yet relatively flexible weapon systems based on the basic design principle of 'reliability' (sensu Bleed, 1986). The primary functional constraint would have been their ability to penetrate thick-skinned animal species such as the European moose or Giant Elk (cf. Bokelmann, 1978; Tomka, 2013), while ballistic considerations generally suggest that large tanged points serving as dart tips should have had relatively lax constraints on size and weight (cf. Riede, 2009). The whole-shape variability recorded in our Final Palaeolithic/earliest Mesolithic tanged points dataset generally supports this interpretation, indicating that especially the distal part of these points and, thus their 'business ends' (Kuhn, 2020), allowed for considerable formal variability and that normalising this part of the tool was not a primary concern; design imperatives rather ensured their convergence or pointedness and not so much particular morphological or technical details connected to point tip and lateral edge configuration. The consistency of the tanged element, often prepared in more elaborate fashion, is arguably the central design feature of these points and in fact suggests that another source of constraint was at least equally important: hafting. The coherent elaboration of a structurally differentiated proximal part would support what Kuhn (2020, 46) terms 'Mode B' hafting — the strategy of custom-fitting a stone object to a durable handle or intermediate part made of another type of raw material (cf. Barham, 2013). The fact that various reduction approaches and modification tactics were apparently employed to achieve essentially the same outcome in the European Final Palaeolithic — a tang-like

prehensive tool part — substantiates this conclusion, pointing towards hafting requirements as a key dimension of the overall design of these large tanged points.

It is salubrious to keep in mind that prehistoric stone artefacts are non-industrial products (Kuhn, 2020; Shea, 2017) and that all knapping has an unescapably experimental quality (Weißmüller, 1995, 2003) which leads to much natural variability or ‘noise’ in the data, especially if this variability is not deliberately removed by extensive secondary modification. This combination of limited functional constraints, influence of the mode of production and copying error in cultural transmission (cf. Eerkens & Lipo, 2007) as well as other processes such as horizontal transmission and exchange of weapons or weapon tips (Nishiaki, 2013; Wiessner, 1983) may have combined to obscure any cultural phylogenetic signal in this particular artefact class. Whether the signal is further compounded by a combination of low population density and the multifarious impacts of the Laacher See eruption (see Riede, 2017) is up to debate. An alternative interpretation of the absence of any cultural phylogenetic signal in this dataset — apart from the concerns about chronological resolution and sample size raised above — may thus be methodological: no cultural evolutionary information may be inherent in outline shapes because outline is only indirectly and often weakly controlled by potentially learned and transmitted tool-making behaviours.

Whatever the ultimate reasons, the internal sum of variance at least demonstrates that the assemblage of large tanged points studied here and falling into a time frame of approximately 2000–3000 years is characterised by considerably lower combined outline disparity than the post-Palaeolithic arrowheads with a much narrower spatiotemporal spread (Table 3). One possibility to interpret this difference is to regard the sample of Final Palaeolithic/earliest Mesolithic tanged points as an expression of slow-paced, ‘thick time’ (sensu Shryock et al., 2012) temporalities of change, resulting in much formal supra-regional homogeneity and markedly reduced rates of morphological evolution. Again, this might be linked to the firm integration of these large tanged points into modular weapon systems, while the present sample of post-Palaeolithic points may exhibit functional diversification and specialisation (cf. Nicolas, 2016) mediated by a sociocultural context in which tool status was highly dynamic and possibly responsive to social and economic stratification. This interpretation is also consistent with the data obtained from the Paleoindian sample, yielding the lowest total sum of variance of all studied datasets, even though its chronological and spatial coverage is substantial (Table 3). Paleoindian point designs are generally consistent with principles of ‘reliability’ as originally outlined by Bleed (1986), whereas European later prehistoric points seem to combine aspects of reliability with deliberate stylistic variation driven by concerns of status and identity — what Wiessner (1983) has influentially termed ‘emblematic style’. The Final Palaeolithic large tanged points exhibit both overall shape variability — design concerns apparently centre on the creation of structurally differentiated prehensile parts — and little in the way of internal spatiotemporal patterning suggesting generally expedient point designs and largely ‘embedded’ or ‘isochrestic’ style regimes (sensu Sackett, 1982). All of these cases reflect substantially different design solutions, and careful comparative whole-outline analysis is able to make them evident.

Table 3 Summary of the analysed datasets, sample characteristics and main results including calculated disparity values

Dataset	Geographic extent	Temporal extent	Archaeological context	N	Methods applied	Findings	Disparity
Paleoindian Goshen and Plainview projectile points (Buchanan et al., 2020)	Northern Great Plains (US)	12,500–11300 cal BC (1200 years)	Stratified archaeological sites	156	PERMANOVA, LDA	Statistical tests on their outline shape cannot distinguish between the two types	0.020
Bell Beaker projectile points (Peřfík et al., 2018)	Moravia (Czech Republic)	2500–2300/2200 BC (200–300 years)	~75% graves, ~15% excavation (other), ~10% surface collection	189	Hierarchical clustering (UPGMA)	Similar clusters as in the landmark-based study were recovered	0.025
Late Neolithic/Early Bronze Age arrow heads (Nicolas, 2016)	Southern England, north-western France, Denmark	2500–1700 cal BC (800 years)	~90% single graves, ~1.5% double graves, ~7.5% collective burials	783	Hierarchical clustering (Ward's method), inference of phenogram (NJ algorithm)	The artefacts were separated into sensible clusters according to their form and geographical spread. The original author's typo-chronology for the French specimen could be reconstructed using their outline data in combination with the NJ algorithm	0.078 (combined); 0.047 for FR and the UK, respectively, 0.042 for DK

Table 3 (continued)

Dataset	Geographic extent	Temporal extent	Archaeological context	N	Methods applied	Findings	Disparity
Final Palaeolithic/ earliest Mesolithic large tanged points (Riede et al., 2019)	France, Belgium, the Netherlands, Ger- many, Denmark, Sweden, Poland, Belarus, Ukraine, Russia, Lithuania	Ranging from Bølling–Preboreal (ca. 14–10.7 ka cal BP; ~3000 years), with most artefacts falling into the Allerød–Younger Dryas period (ca. 13.6–11.6 ka cal BP; ~2000 years)	~48.5% excavation, ~50% surface collection, ~1.5% mixed	253	Hierarchical clus- tering (Ward's method), inference of phenogram (NJ algorithm)	No diagnostic pat- terns in form, space and time could be recovered	0.043

In conclusion, our findings have important implications — both methodological and theoretical — for better understanding and explaining the relationship between cultural evolutionary processes and shape dynamics in lithic points and also for how to contribute to and advance the broader debate on style, function and design underwriting processes of stone tool manufacture and use in archaeology and beyond. 2D whole-outline analyses as developed here could, for instance, be used to test long-standing ideas about the changing status of shape-anchored lithic types through time (cf. Weißmüller, 2003; Valentin, 2011). Future studies may build on the methods applied here and, if possible, explore if and when rates of shape evolution experienced major changes and when and how design spaces were subjected to reconfiguration, ideally comparing multiple periods, transitional phases and regions. In palaeobiology, new computational methods for inferring phylogenies from continuous data are currently emerging (Parins-Fukuchi, 2017; 2018) and may yet offer a powerful extension of the methodology presented here. In particular, these approaches are adopting Bayesian inference methods (Wright, 2019; Warnock & Wright, 2020) and offer an integration with the downstream analytical tools of the so-called ‘comparative method’ (e.g. Adams & Collyer, 2019) which has seen only limited use in Palaeolithic archaeology (cf. Riede, 2011). Alternatively, a discrete character-based phylogenetic approach could be implemented and tested, systematically mapping a range of multivariate traits, such as tool outlines and contextual ecological variables.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s41982-021-00097-2>.

Acknowledgements The authors would like to express their gratitude to Briggs Buchanan and Jan Petřík for providing original photographs and drawings and their agreement to reproduce some of the original images here. We thank two anonymous reviewers and the editors of this journal for constructive comments and suggestions which have helped to improve the original manuscript. All remaining shortcomings are our own responsibility.

Author contributions Conceptualisation, Felix Riede, David N. Matzig and Shumon T. Hussain; methodology, Felix Riede and David N. Matzig; software, David N. Matzig; formal analysis and investigation, David N. Matzig; data curation, David N. Matzig; writing — original draft preparation, Felix Riede, David N. Matzig and Shumon T. Hussain; writing — review and editing, Felix Riede, David N. Matzig and Shumon T. Hussain; visualisation, David N. Matzig; funding acquisition, Felix Riede; resources: Felix Riede; supervision: Felix Riede and Shumon T. Hussain.

Funding This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 817564) as well as the Independent Research Fund Denmark (#6107-00059B).

Data availability The datasets generated and/or analysed in the current study are available in the associated research compendium available on Zenodo: <https://doi.org/10.5281/zenodo.4560743>.

Code Availability All analyses and visualisations presented in this paper were prepared in R 3.6.3 (R Core, 2020) on Ubuntu 18.04.5 LTS (64-bit) using the following packages: *ape* (>= 5.4-1), *asbio* (>= 1.6-5), *BiocManager* (>= 1.30.10), *caret* (>= 6.0-86), *cluster* (>= 2.1.0), *coin* (>= 1.3-1), *cowplot* (>= 1.0.0), *disprity* (>= 1.5.1), *dplyr* (>= 1.0.0.9000), *fpc* (>= 2.2-7), *ggimage* (>= 0.2.8), *ggplot2* (>= 3.3.2), *ggpubr* (>= 0.4.0), *ggthemes* (>= 4.2.0), *ggtree* (>= 2.0.4), *maptree* (>= 1.4-7), *Momocs* (>= 1.3.0), *MVN* (>= 5.8), *NbClust* (>= 3.0), *parallel* (>= 3.6.3), *phangorn* (>= 2.5.5), *psych* (>= 1.9.12.31), *randomcolor* (>= 1.1.0.1), *raster* (>= 3.3-7), *readr* (>= 1.3.1), *rgeos* (>= 0.5-3), *rworldmap*

(>= 1.3-6), *splitstackshape* (>= 1.4.8), *vegan* (>= 2.5-6). All code associated with this paper is available in the accompanying research compendium on Zenodo under <https://doi.org/10.5281/zenodo.4560743>.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adams, D. C., & Collyer, M. L. (2019). Phylogenetic comparative methods and the evolution of multivariate phenotypes. *Annual Review of Ecology, Evolution, and Systematics*, 50(1), 405–425. <https://doi.org/10.1146/annurev-ecolsys-110218-024555>
- Albrecht, G., & Müller-Beck, H. (1988). The Palaeolithic of Sehremuz near Samsat on the Euphrates River. Summary of the excavation findings and a morphology of the handaxes. *Paléorient*, 14(2), 76–86. <https://doi.org/10.3406/paleo.1988.4457>
- Barham, L. (2013). *From hand to handle: The first industrial revolution*. Oxford University Press.
- Barthelme, S., Tschumperle, D., Wijffels, J., Assemlal, H. E., & Ochi, S. (2020). Imager: Image processing library based on 'CImg'.
- Birch, T., & Martínón-Torres, M. (2019). Shape as a measure of weapon standardisation: From metric to geometric morphometric analysis of the Iron Age “Havor” lance from Southern Scandinavia. *Journal of Archaeological Science*, 101, 34–51. <https://doi.org/10.1016/j.jas.2018.11.002>
- Bisson, M. S. (2000). Nineteenth century tools for twenty-first century archaeology? Why the Middle Paleolithic typology of François Bordes must be replaced. *Journal of Archaeological Method and Theory*, 7(1), 1–48. <https://doi.org/10.1023/A:1009578011590>
- Bleed, P. (1986). The optimal design of hunting weapons: Maintainability or reliability. *American Antiquity*, 51(4), 737–747.
- Boëda, E. (2013). *Techno-logique & Technologie. Une Paléo-histoire des objets lithiques tranchants*. Archéo-éditions.
- Bokelmann, K. (1978). Ein Federmesserfundplatz bei Schalkholz, Kreis Dithmarschen. *Offa. Berichte und Mitteilungen zur Urgeschichte, Frühgeschichte und Mittelalterarchäologie*, 35, 36–54.
- Bonhomme, V., Picq, S., Gauchere, C., & Claude, J. (2014). Momocs: Outline analysis using R. 56(13): 1–24. <https://doi.org/10.18637/jss.v056.i13>.
- Buchanan, B., Collard, M., & O'Brien, M. J. (2020). Geometric morphometric analyses support incorporating the Goshen point type into plainview. *American Antiquity*, 85(1), 171–181. <https://doi.org/10.1017/aaq.2019.89>
- Burdukiewicz, J. M. (1986). *Late Pleistocene shouldered point assemblages in Western Europe*. E.J. Brill.
- Cardillo, M. (2010). Some applications of geometric morphometrics to archaeology. In A. M. T. Elewa (Ed.), *Morphometrics for Nonmorphometricians* (pp. 325–341). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-95853-6_15.
- Chacón, M. G., Détroit, F., Coudenneau, A., & Moncel, M.-H. (2016). Morphometric assessment of convergent tool technology and function during the Early Middle Palaeolithic: The case of Payre, France. *PLoS One*, 11(5), e0155316. <https://doi.org/10.1371/journal.pone.0155316>

- Charlin, J., & González-José, R. (2012). Size and shape variation in Late Holocene projectile points of Southern Patagonia: A geometric morphometric study. *American Antiquity*, 77(2), 221–242. <https://doi.org/10.7183/0002-7316.77.2.221>
- Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data Set. *Journal of Statistical Software*, 61(1), 1–36. <https://doi.org/10.18637/jss.v061.i06>
- Clark, G. A., & Riel-Salvatore, J. (2006). Observations on systematics in Paleolithic archaeology. In E. Hovers & S. L. Kuhn (Eds.), *Transitions Before the Transition: Evolution and Stability in the Middle Paleolithic and Middle Stone Age* (pp. 29–56). Springer, US.
- Clarke, D. L. (1968). *Analytical archaeology* (1st ed.). Methuen & Co Ltd.
- Cooke, S. B., & Terhune, C. E. (2015). Form, function, and geometric morphometrics. *The Anatomical Record*, 298(1), 5–28. <https://doi.org/10.1002/ar.23065>
- Costa, A. G. (2010). A geometric morphometric assessment of plan shape in bone and stone Acheulean bifaces from the Middle Pleistocene site of Castel di Guido, Latium, Italy. In S. Lycett & P. Chauhan (Eds.), *New Perspectives on Old Stones* (pp. 23–41). Springer.
- Dietrich, L., Rokitta-Krumnow, D., & Dietrich, O. (2019). The meaning of projectile points in the Late Neolithic of the Northern Levant. *Documenta Praehistorica*, XLVI, 340–350. <https://doi.org/10.4312/dp.46.21>
- Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, 14(6), 927–930. <https://doi.org/10.1111/j.1654-1103.2003.tb02228.x>
- Eerkens, J. W., & Lipo, C. P. (2007). Cultural transmission theory and the archaeological record: Providing context to understanding variation and temporal changes in material culture. *Journal of Archaeological Research*, 15(3), 239–274. <https://doi.org/10.1007/s10814-007-9013-z>
- Felsenstein, J. (2004). *Inferring phylogenies*. Sinauer Associates.
- Florek, J. K. (1951). Sur la liaison et la division des points d'un ensemble fini. *Colloquium Mathematicum*, 2(3–4), 282–285.
- Fox, A. N. (2015). A study of Late Woodland projectile point typology in New York using elliptical Fourier outline analysis. *Journal of Archaeological Science: Reports*, 4, 501–509. <https://doi.org/10.1016/j.jasrep.2015.10.022>
- Fugl Petersen, B. (2001). Senpaleolitiske opsamlingsfund fra Sydsjælland, Fejø og Nordsjælland - et bidrag til udforskningen af de senglaciale kulturer i Danmark. *Kulturhistoriske Studier*, 2001, 7–64.
- Gamble, C., Davies, W., Pettitt, P., Hazelwood, L., & Richards, M. (2005). The archaeological and genetic foundations of the European population during the Late Glacial: Implications for “agricultural thinking.” *Cambridge Archaeological Journal*, 15(2), 193–223. <https://doi.org/10.1017/S0959774305000107>
- Geupel, V. (1987). *Spätpaläolithikum und Mesolithikum im Süden der DDR: Katalog Teil 2*. VEB Deutscher Verlag der Wissenschaften.
- González-José, R., Escapa, I., Neves, W. A., Cúneo, R., & Pucciarelli, H. M. (2008). Cladistic analysis of continuous modularized traits provides phylogenetic signals in Homo evolution. *Nature*, 453(7196), 775–778. <https://doi.org/10.1038/nature06891>
- Gopher, A. (1994). *Arrowheads of the Neolithic Levant: A seriation analysis*. (B. Halpern, Ed.) (vol. 10). Eisenbrauns.
- Gowlett, J. A. J. (2006). The elements of design form in Acheulean bifaces: Modes, modalities, rules, and language. In N. Goren-Inbar & G. Sharon (Eds.), *Axe Age: Acheulean Tool-making from Quarry to Discard* (pp. 203–221). Equinox.
- Guillierme, T. (2018). dispRity : A modular R package for measuring disparity. *Methods in Ecology and Evolution*, 9(7), 1755–1763. <https://doi.org/10.1111/2041-210X.13022>
- Hull, D. L. (1965). The effect of essentialism on taxonomy - Two thousand years of stasis. *British Journal for the Philosophy of Science*, 15, 314–326.
- Iovita, R. (2009). Ontogenetic scaling and lithic systematics: Method and application. *Journal of Archaeological Science*, 36(7), 1447–1457.
- Ivanovaitė, L., Serwatka, K., Hoggard, C. S., Sauer, F., & Riede, F. (2020). All these fantastic cultures? Research history and regionalization in the Late Palaeolithic tanged point cultures of Eastern Europe. *European Journal of Archaeology*, 23(2), 162–185. <https://doi.org/10.1017/eea.2019.59>
- Klingenberg, C. P., & Monteiro, L. R. (2005). Distances and directions in multidimensional shape spaces: Implications for morphometric applications. *Systematic Biology*, 54(4), 678–688. <https://doi.org/10.1080/10635150590947258>

- Kopacz, J., Přichystal, A., & Šebela, L. (2009). *Lithic chipped industry of the Bell Beaker culture in Moravia and its East-Central European context*. Polska Akad. Umiejętności
- Kuhl, F. P., & Giardina, C. R. (1982). Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing*, 18(3), 236–258. [https://doi.org/10.1016/0146-664X\(82\)90034-X](https://doi.org/10.1016/0146-664X(82)90034-X)
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., & et al. (2020). Caret: Classification and regression training.
- Kuhn, S. L. (2020). *The evolution of Paleolithic technologies*. Routledge.
- Leplongeon, A., Ménard, C., Bonhomme, V., & Bortolini, E. (2020). Backed pieces and their variability in the Later Stone Age of the Horn of Africa. *African Archaeological Review*. <https://doi.org/10.1007/s10437-020-09401-x>
- Lepot, M. (1992). *Approche techno-fonctionnelle de l'outillage lithique moustérien : Essai de classification des parties actives en termes d'efficacité technique. Application à la couche M2e sagittale du Grand Abri de La Ferrassie (fouille Henri Delporte)*. Université de Paris X Nanterre: Unpublished Mémoire de la Maîtrise.
- Lipo, C. P., Hunt, T. L., Horneman, R., & Bonhomme, V. (2016). Weapons of war? Rapa Nui mata'a morphometric analyses. *Antiquity*, 90(349), 172–187. <https://doi.org/10.15184/aqy.2015.189>
- Lipo, C. P., O'Brien, M. J., Collard, M., & Shennan, S. J. (Eds.). (2006). *Mapping our ancestors. Phylogenetic approaches in anthropology and prehistory*. Aldine Transaction.
- Lycett, S. J. (2007). Is the Soanian techno-complex a Mode 1 or Mode 3 phenomenon? A morphometric assessment. *Journal of Archaeological Science*, 34(9), 1434–1440.
- Lycett, S. J. (2009). Quantifying transitions: Morphometric approaches to Palaeolithic variability and technological change. In M. Camps, & P. Chauhan (Eds.), *Sourcebook of Paleolithic Transitions: Methods, Theories, and Interpretations* (pp. 79–92). Springer New York. https://doi.org/10.1007/978-0-387-76487-0_5.
- Lycett, S. J. (2010). The importance of history in definitions of culture: Implications from phylogenetic approaches to the study of social learning in chimpanzees. *Learning & Behavior*, 38(3), 252–264. <https://doi.org/10.3758/lb.38.3.252>
- Mace, R., Holden, C. J., & Shennan, S. J. (2005). *The evolution of cultural diversity. A Phylogenetic Approach*. UCL Press.
- Mace, R., & Pagel, M. D. (1997). Phylogenies and cultural evolution. *Evolution and Human Behavior*, 18(5), 349–351.
- MacLeod, N. (1999). Generalizing and extending the eigenshape method of shape space visualization and analysis. *Paleobiology*, 25(1), 107–138.
- MacLeod, N. (2001). Landmarks in phylogenetic analysis. In J. M. Adrain, G. D. Edgecombe, & B. S. Lieberman (Eds.), *Fossils, phylogeny, and form: An analytical approach* (pp. 197–233). Kluwer Academic/Plenum Publishers.
- MacLeod, N. (2018). The quantitative assessment of archaeological artifact groups: Beyond geometric morphometrics. *Quaternary Science Reviews*, 201, 319–348. <https://doi.org/10.1016/j.quascirev.2018.08.024>
- Mahto, A. (2019). *Splitstackshape: Stack and reshape datasets after splitting concatenated values*. Manual.
- Matzig, D. N. (2021). outlineR: An R package to derive outline shapes from (multiple) artefacts on JPEG images. Zenodo. <https://doi.org/10.5281/ZENODO.4527469>.
- Mayr, E. (1959). Typological versus population thinking. In B. J. Meggers (Ed.), *Evolution and anthropology: A centennial appraisal* (pp. 409–412). The Anthropological Society of Washington.
- McNabb, J., Cole, J., & Hoggard, C. S. (2018). From side to side: Symmetry in handaxes in the British Lower and Middle Palaeolithic. *Journal of Archaeological Science: Reports*, 17, 293–310. <https://doi.org/10.1016/j.jasrep.2017.11.008>
- Mesfin, I., Leplongeon, A., Pleurdeau, D., & Borel, A. (2020). Using morphometrics to reappraise old collections: The study case of the Congo Basin Middle Stone Age bifacial industry. *Journal of Lithic Studies*, 7(1). <https://doi.org/10.2218/jls.4329>.
- Migal, W. (2007). On preferential points of the Final Paleolithic in the Central European Lowland. In M. Kobusiewicz, & J. Kabacinski (Eds.), *Studies in the Final Palaeolithic settlement of the Great European Plain* (pp. 185–200). Institute of Archaeology and Ethnology, Polish Academy of Sciences (Poznan Branch).
- Monnier, G. F., & McNulty, K. P. (2010). Questioning the link between stone tool standardization and behavioral modernity. In S. Lycett, & P. Chauhan (Eds.), *New Perspectives on Old Stones* (pp. 61–81). Springer New York. https://doi.org/10.1007/978-1-4419-6861-6_4.

- Monnier, G. F., & Missal, K. (2014). Another Mousterian debate? Bordian facies, chaîne opératoire technocomplexes, and patterns of lithic variability in the western European Middle and Upper Pleistocene. *Quaternary International*, 350, 59–83. <https://doi.org/10.1016/j.quaint.2014.06.053>
- Nakoinz, O. (2013). *Archäologische Kulturgeographie der ältereisenzeitlichen Zentralorte Südwestdeutschlands*. Habelt.
- Nicolas, C. (2016). *Flèches de pouvoir à l'aube de la métallurgie de la Bretagne au Danemark (2500-1700 av. n.è.)*. Sidestone Press.
- Nicolas, C. (2017). Arrows of power from Brittany to Denmark (2500 BC). *Proceedings of the Prehistoric Society*, 83, 247–287. <https://doi.org/10.1017/ppr.2017.5>
- Nishiaki, Y. (2013). “Gifting” as a means of cultural transmission: The archaeological implications of bow-and-arrow technology in Papua New Guinea. In T. Akazawa, Y. Nishiaki, & K. Aoki (Eds.), *Dynamics of Learning in Neanderthals and Modern Humans* (pp. 173–185). Springer.
- O’Brien, M. J., Buchanan, B., & Eren, M. I. (2018). *Convergent evolution in stone-tool technology*. The MIT Press.
- O’Brien, M. J., & Lyman, R. L. (2002). The epistemological nature of archaeological units. *Anthropological Theory*, 2(1), 37–56. <https://doi.org/10.1177/1463499602002001287>
- O’Brien, M. J., Lyman, R. L., Collard, M., Holden, C. J., & Shennan, S. J. (2008). Transmission, phylogenetics, and the evolution of cultural diversity. In M. J. O’Brien (Ed.), *Cultural Transmission in Archaeology: Issues and Case Studies* (pp. 39–58). Society for American Archaeology Press.
- Okumura, M., & Araujo, A. G. M. (2014). Long-term cultural stability in hunter gatherers: A case study using traditional and geometric morphometric analysis of lithic stemmed bifacial points from Southern Brazil. *Journal of Archaeological Science*, 45, 59–71. <https://doi.org/10.1016/j.jas.2014.02.009>
- Okumura, M., & Araujo, A. G. M. (2018). Archaeology, biology, and borrowing: A critical examination of geometric morphometrics in archaeology. *Journal of Archaeological Science*. <https://doi.org/10.1016/j.jas.2017.09.015>
- Olivík, J. (2009). *Silicovitové šípky a nátepní destičky - jejich typologie a postavení v pohřebním ritu kultury zvoncovitých pohárů na moravě* (Diplomová práce). Masaryk University, Brno. Retrieved from <https://is.muni.cz/th/ejukm/> (05/05/2021).
- Parins-Fukuchi, C. (2017). Use of continuous traits can improve morphological phylogenetics. *Systematic Biology*, 67(2), 328–339. <https://doi.org/10.1093/sysbio/syx072>
- Parins-Fukuchi, C. (2018). Bayesian placement of fossils on phylogenies using quantitative morphometric data. *Evolution*, 72(9), 1801–1814. <https://doi.org/10.1111/evo.13516>
- Pau, G., Fuchs, F., Sklyar, O., Boutros, M., & Huber, W. (2010). EBImagean R package for image processing with applications to cellular phenotypes. *Bioinformatics*, 26(7), 979–981. <https://doi.org/10.1093/bioinformatics/btq046>
- Pesesse, D., & Flas, D. (2012). The Maisierian, at the Edge of the Gravettian. *Proceedings of the Prehistoric Society*, 78, 95–109. <https://doi.org/10.1017/S0079497X00027122>
- Petrík, J., Sosna, D., Prokeš, L., Štefanisko, D., & Galeta, P. (2018). Shape matters: Assessing regional variation of Bell Beaker projectile points in Central Europe using geometric morphometrics. *Archaeological and Anthropological Sciences*, 10(4), 893–904. <https://doi.org/10.1007/s12520-016-0423-z>
- Polly, P. D. (2001). Paleontology and the comparative method: Ancestral node reconstructions versus observed node values. *American Naturalist*, 157(6), 596–609. <https://doi.org/10.1086/320622>
- R Core Team. (2020). R: A Language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Reyment, R. A. (1985). Multivariate morphometrics and analysis of shape. *Journal of the International Association for Mathematical Geology*, 17(6), 591–609. <https://doi.org/10.1007/BF01030855>
- Reynolds, N., & Riede, F. (2019). House of cards: Cultural taxonomy and the study of the European Upper Palaeolithic. *Antiquity*, 93(371), 1350–1358. <https://doi.org/10.15184/aqy.2019.49>
- Richter, J. (2013). Bewusste Geometrische Gestaltung bei Homo Heidelbergensis? Arbeitsschrittanalyse an einem Faustkeil aus Bad Salzuffen (Ostwestfalen-Lippe). *Archäologisches Korrespondenzblatt*, 43(1), 1–17.
- Riede, F. (2009). The loss and re-introduction of bow-and-arrow technology: A case study from the Northern European Late Paleolithic. *Lithic Technology*, 34(1), 27–45. <https://doi.org/10.1016/j.aanat.2009.12.003>

- Riede, F. (2011). Adaptation and niche construction in human prehistory: A case study from the southern Scandinavian Late Glacial. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1566), 793–808. <https://doi.org/10.1098/rstb.2010.0266>
- Riede, F. (2017). *Splendid isolation. The eruption of the Laacher See volcano and southern Scandinavian Late Glacial hunter-gatherers*. Aarhus University Press.
- Riede, F., Araujo, A. G. M., Barton, M. C., Bergsvik, K. A., Groucutt, H. S., Hussain, S. T., et al. (2020). Cultural taxonomies in the Paleolithic Old questions, novel perspectives. *Evolutionary Anthropology: Issues, News, and Reviews*, 29(2), 49–52. <https://doi.org/10.1002/evan.21819>
- Riede, F., Hoggard, C., & Shennan, S. (2019). Reconciling material cultures in archaeology with genetic data requires robust cultural evolutionary taxonomies. *Palgrave Communications*, 5(1). <https://doi.org/10.1057/s41599-019-0260-7>.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- Sackett, J. R. (1982). Approaches to style in lithic archaeology. *Journal of Anthropological Archaeology*, 1, 59–112.
- Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>
- Sarauw, T. (2008). Danish Bell Beaker pottery and flint daggers - The display of social identities? *European Journal of Archaeology*, 11(1), 23–47. <https://doi.org/10.1177/1461957108101240>
- Schild, R., Pazdur, A., & Vogel, J. C. (1999). Radiochronology of the tanged point technocomplex in Poland. In S. K. Kozłowski, J. Gurba, & L. L. Zaliznyak (Eds.), *Tanged Point Cultures in Europe. Read at the International Archaeological Symposium. Lublin, September, 13-16, 1993* (Vol. Lubelskie Materiały Archeologiczne 13, pp. 13–15). Maria Curie-Skłodowska University Press.
- Schliep, K. P. (2010). Phangorn: Phylogenetic analysis in R. *Bioinformatics*, 27(4), 592–593. <https://doi.org/10.1093/bioinformatics/btq706>
- Seong, C. (2008). Tanged points, microblades and Late Palaeolithic hunting in Korea. *Antiquity*, 82, 871–883.
- Serwatka, K. (2014). Shape variation of Middle Palaeolithic bifacial tools from southern Poland: A geometric morphometric approach to *Keilmessergruppen* handaxes and backed knives. *Lithics*, 35, 18–32.
- Serwatka, K., & Riede, F. (2016). 2D geometric morphometric analysis casts doubt on the validity of large tanged points as cultural markers in the European Final Palaeolithic. *Journal of Archaeological Science: Reports*, 9, 150–159. <https://doi.org/10.1016/j.jasrep.2016.07.018>
- Shea, J. J. (2014). Sink the Mousterian? Named stone tool industries (NASTIES) as obstacles to investigating hominin evolutionary relationships in the Later Middle Paleolithic Levant. *Quaternary International*, 350, 169–179. <https://doi.org/10.1016/j.quaint.2014.01.024>
- Shea, J. J. (2017). Occasional, obligatory, and habitual stone tool use in hominin evolution. *Evolutionary Anthropology: Issues, News, and Reviews*, 26(5), 200–217. <https://doi.org/10.1002/evan.21547>
- Shott, M. J., & Trail, B. W. (2010). Exploring new approaches to lithic analysis: Laser scanning and geometric morphometrics. *Lithic Technology*, 35(2), 195–220.
- Shryock, A., Smail, D. L., & Earle, T. K. (Eds.). (2012). *Deep history: The architecture of past and present*. Berkeley, Calif.: Univ. of California Press.
- Simpson, G. G. (1945). The principles of classification and a classification of mammals. *Bulletin of the American Museum of Natural History*, 85, 1–350.
- Smith, U. E., & Hendricks, J. R. (2013). Geometric morphometric character suites as phylogenetic data: Extracting phylogenetic signal from gastropod shells. *Systematic Biology*, 62(3), 366–385. <https://doi.org/10.1093/sysbio/syt002>
- Sober, E. (1980). Evolution, population thinking, and essentialism. *Philosophy of Science*, 47(3), 350–383.
- Sokal, R. R., & Michener, C. D. (1958). *A statistical method for evaluating systematic relationships*. University of Kansas.
- Sørensen, T. J. (1948). *A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons*. I kommission hos E. Munksgaard.

- Studier, J. A., & Keppler, K. J. (1988). A note on the neighbor-joining algorithm of Saitou and Nei. *Molecular Biology and Evolution*, 5(6), 729–731. <https://doi.org/10.1093/oxfordjournals.molbev.a040527>
- Suárez, R., & Cardillo, M. (2019). Life history or stylistic variation? A geometric morphometric method for evaluation of Fishtail point variability. *Journal of Archaeological Science: Reports*, 27, 101997. <https://doi.org/10.1016/j.jasrep.2019.101997>
- Temple, J. T. (1992). The progress of quantitative methods in palaeontology. *Palaeontology*, 35(2), 475–484.
- Tomka, S. A. (2013). The adoption of the bow and arrow: A model based on experimental performance characteristics. *American Antiquity*, 78(3), 553–569.
- Tostevin, G. B. (2019). Content matters: The materiality of cultural transmission and the intersection of Paleolithic archaeology with cultural evolutionary theory. In A. C. Love, & W. C. Wimsatt (Eds.), *Beyond the Meme: Development and Structure in Cultural Evolution* (pp. 311–364). University of Minnesota Press. <https://doi.org/10.5749/j.ctvnp0krm>.
- Valentin, B. (2011). *Le paléolithique*. Presses universitaires de France.
- Venables, W. N., & Ripley, B. D. (2002). *BD Ripley 2002 Modern applied statistics with s (Fourth)*. Springer.
- Vestergaard, C., & Hoggard, C. S. (2019). A novel geometric morphometric (GMM) application to the study of Bronze Age Tutuli. *Danish Journal of Archaeology*, 8, 1–24. <https://doi.org/10.7146/dja.v8i0.112494>
- Wang, L.-Y., & Marwick, B. (2020). Standardization of ceramic shape: A case study of Iron Age pottery from northeastern Taiwan. *Journal of Archaeological Science: Reports*, 33, 102554. <https://doi.org/10.1016/j.jasrep.2020.102554>
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236–244. <https://doi.org/10.1080/01621459.1963.10500845>
- Warnock, R. C. M., & Wright, A. M. (2020). *Understanding the tripartite approach to Bayesian divergence time estimation* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/9781108954365>.
- Weißmüller, W. (1995). *Die Silexartefakte der Unteren Schichten der Sesselfelsgrötte. Ein Beitrag zum Problem des Moustérien* (Vol. 6). Saarbrücken.
- Weißmüller, W. (2003). Von Gestaltungsfreiheit und Standardisierungszwang: Zum Zeichencharakter paläolithischer Silexartefakte. In U. Veit, T. Kienlin, C. Kümmel, & S. Schmidt (Eds.), *Spuren und Botschaften: Interpretationen materieller Kultur* (pp. 173–185). Waxmann.
- Whittaker, J. C. (1994). *Flint knapping: Making and understanding stone tools*. University of Texas Press.
- Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.
- Wiessner, P. (1983). Style and social information in Kalahari San projectile points. *American Antiquity*, 48(2), 253–276. <https://doi.org/10.2307/280450>
- Wilczek, J., Monna, F., Gabillot, M., Navarro, N., Rusch, L., & Chateau, C. (2015). Unsupervised model-based clustering for typological classification of Middle Bronze Age flanged axes. *Journal of Archaeological Science: Reports*, 3, 381–391. <https://doi.org/10.1016/j.jasrep.2015.06.030>
- Wilkins, J. (2020). Is it time to retire NASTIES in Southern African? Moving beyond the culture-historical framework for Middle Stone Age Lithic Assemblage Variability. *Lithic Technology*, 1–13. <https://doi.org/10.1080/01977261.2020.1802848>.
- Williams, T. J., & Madsen, D. B. (2019). The Upper Paleolithic of the Americas. *PaleoAmerica*, 1–19. <https://doi.org/10.1080/20555563.2019.1606668>.
- Winkler, K. (2010). Der spätpaläolithische Fundplatz Golßen, Lkr. Dahme-Spreewald. Analyse der Steinartefakte und Befunde der Grabung von 1968. Veröffentlichungen zur brandenburgischen Landesarchäologie, 41/42, 7–93.
- Witteveen, J. (2015). “A temporary oversimplification”: Mayr, Simpson, Dobzhansky, and the origins of the typology/population dichotomy (part 1 of 2). *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 54, 20–33. <https://doi.org/10.1016/j.shpsc.2015.09.007>
- Witteveen, J. (2018). Typological thinking: Then and now. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 330(3), 123–131. <https://doi.org/10.1002/jez.b.22796>
- Wright, A. M. (2019). A Systematist’s guide to estimating Bayesian phylogenies from morphological data. *Insect Systematics and Diversity*, 3(3): 2, 1–14. <https://doi.org/10.1093/isd/ixz006>.

- Zelditch, M. L., Fink, W. L., & Swiderski, D. L. (1995). Morphometrics, homology, and phylogenetics: Quantified characters as synapomorphies. *Systematic Biology*, 44(2), 179. <https://doi.org/10.2307/2413705>
- Zelditch, M. L., Swiderski, D. L., & Fink, W. L. (2001). Homology, characters, and morphometric data. In J. M. Adrain, G. D. Edgecombe, & B. S. Lieberman (Eds.), *Fossils, phylogeny, and form: An analytical approach* (pp. 145–196). Kluwer Academic/Plenum Publishers.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

David N. Matzig¹ · Shumon T. Hussain¹ · Felix Riede¹

✉ David N. Matzig
david.matzig@cas.au.dk

¹ Department of Archaeology and Heritage Studies, Aarhus University, Aarhus, Denmark