The Potential to Infer the Historical Pattern of Cultural Macroevolution

Dieter Lukas^{1*}, Mary Towner², Monique Borgerhoff Mulder^{1,3}

- 1) Department of Human Behavior, Ecology & Culture, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany; * author for correspondence (dieter_lukas@eva.mpg.de).
- 2) Department of Integrative Biology, Oklahoma State University, Stillwater, OK 74078, USA
- 3) Department of Human Behavior, Ecology & Culture, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany; Department of Anthropology, University of California Davis, Davis, CA 95616, USA

Summary

Phylogenetic analyses increasingly take centre-stage in our understanding of the processes shaping patterns of cultural diversity and cultural evolution over time. Just as biologists explain the origins and maintenance of trait differences among organisms using phylogenetic methods, so anthropologists studying cultural macroevolutionary processes use phylogenetic methods to uncover the history of human populations and the dynamics of culturally transmitted traits. In this paper we revisit concerns with the validity of these methods. Specifically, we use simulations to reveal how properties of the sample (size, missing data), properties of the tree (shape), and properties of the traits (rate of change, number of variants, transmission mode) might influence the inferences that can be drawn about trait distributions across a given phylogeny and the power to discern alternative histories. Our approach shows that in two example datasets specific combinations of properties of the sample, of the tree, and of the trait can lead to potentially high rates of Type I and Type II errors. We offer this simulation tool to help assess the potential impact of this list of persistent perils in future cultural macroevolutionary work.

1. Introduction

Human societies exhibit a diversity of cultural practices around the world [e.g., 1]. The field of cultural macroevolution aims to identify the origins of this diversity and the factors shaping the distribution of cultural variation across societies [2]. It is now almost obligatory for studies investigating cross-cultural variation to link cultural practices to phylogenies derived from linguistic, morphological, or genetic data [3-5]. Phylogenetic information is often included in comparative studies, when testing for associations in the distribution of cultural variation, to account for the potential dependencies that arise among traits from a shared history. In addition, to fully understand the history of cultural traits and to determine whether any association among cultural traits does reflect a causal relationship, phylogenetic reconstructions of trait evolution trace changes in cultural practices across a tree reflecting ancestral relationships among the societies. Phylogenetic reconstructions of trait evolution offer the potential to infer what cultural variants might have been present in the past, how variants change from one state to another, and what socio-ecological conditions might have influenced such trajectories of change [6]. As such they allow us to test a wide range of hypotheses for the patterning of human cultural variability, bringing precision to the pursuits of early anthropologists, such as Boas' [7] interest in separating the roles of culture, environment and biology, or Murdock's [8] proposal that changes in residence rules precede change in other social structures.

Phylogenetic reconstruction of trait evolution relies on two steps: the first is to infer the likely historical relationships among populations, and the second is to determine whether

changes in a cultural trait relate to the patterns of historical splits among these populations. Here, we focus on the second part: we assume that a tree is available, and that we want to understand where and when on the tree changes in the cultural traits occurred. In most instances, we do not have information from the past to guide our inferences of the history of a trait. In effect, phylogenetic reconstructions of trait evolution are not opening a window onto the past, but painting a picture about the past based primarily on information from the present.

The accuracy of this picture depends on how well we address at least four challenges. The first is to assess how much the past is likely to resemble the present. If traits change rapidly, we cannot say with certainty which of the variants at the tips of the phylogeny might have been present at a particular point in the past, which in our painting metaphor would be akin to the mix of colours from neighbouring societies leading to brown and blurry internal nodes. Second, the accuracy of the picture is fundamentally affected by the assumptions we use to link the current data to patterns in the past, and whether we use appropriate models of rate and directionality in the transmission of cultural variants. In terms of the metaphor, are we even using the right brush for this type of paint to capture our depiction of the past? These first two challenges question whether a phylogenetic method is appropriate to make inferences about the past for the particular trait. Third, the accuracy will depend on how complete our present information is, on whether we have an adequate sampling of cultural practices to make proper inferences about the past. In terms of the metaphor, do we have the full range of colours in our palette or are some of them missing, and if so why? Fourth, and related, what if there are traits that predominate

in the present that did not exist in the past? We may be working with an entirely inappropriate paint box. The last two challenges relate to the sample available to answer a specific question about the past.

Many studies have investigated the strength of cultural ancestry of various traits by mapping them onto an independent language phylogeny, and then directly evaluate their fit with population history in order to detect phylogenetic signal. For example, Moylan et al [9] examined the distribution of 55 East African cultural traits across a linguistic phylogeny and found that only 18 showed a clear phylogenetic branching pattern. Subsequent studies across multiple cultural domains report widely varying phylogenetic signals for both material [10] and social organizational traits [11], signals that can also vary by scale and the prevalence of cultural boundaries [12, 13]. Even across non-humans, behavioural traits often show very low phylogenetic signal compared to morphological and physiological traits [as in non-human primates, 14] and other animals [15]. This widely varying strength of cultural ancestry likely reflects an interplay of the factors listed above, raising multiple challenges for phylogenetic analyses of cultural traits. These include (a) the extent to which the history of a trait can be reliably reconstructed using population history (as captured in a phylogeny), (b) whether traits change primarily at evolutionary time scales, and (c) whether as investigators we have the appropriate data and (d) samples from which to infer past trait states.

We first examine these four challenges (Section 2) and next present simulations (Section 3) to illustrate some implications of these challenges for inferences in studies of trait

evolution. We end (Section 4) with a discussion of how and why advances in addressing these challenges can make phylogenetic approaches a powerful tool for understanding human cultural diversity. The challenges we review are not fundamental flaws that prohibit cultural phylogenetic approaches, but highlight our need to know the extent to which they affect inferences drawn from phylogenetic analyses. Accordingly, we offer a checklist and our simulation as part of a potential workflow assessing the challenges researchers using phylogenetic methods face, in the spirit of “caveat emptor.”

2. Review of the challenges

(a) To what extent can the history of a trait be reflected by a phylogeny?

(i) Are cultural traits inherited together or do they have independent histories?

Just as biologists recognize that every gene has its own history, so social scientists appreciate this could also be true for cultural traits [e.g., 16, 17]. Boyd et al [18] evaluate how human societies differ with respect to how integrated are their cultural traits. At one end of the continuum, societies are seen as consisting of a tightly-knit set of cultural traits, while at the other end, of largely ephemeral traits with only low coherence. Generally, the empirical evidence points to the middle of the continuum, with societies containing a vertically transmitted core of integrated cultural traits in addition to a hugely varying proportion of more peripheral traits, some horizontally borrowed from other populations and some independently invented. Bayesian phylogenetic analyses can be used to identify incongruent cultural histories, by enabling researchers to classify traits as core or peripheral and then test whether allowing rates of change between partitions provides a better fit with the data than assuming equal rates of change [19]. While this approach can

show how different traits are likely to have had different histories, it cannot reveal the particular history of the individual traits. This leaves unresolved a determination of the extent to which an independently-derived phylogeny can capture the distinct histories of different culturally transmitted traits.

(ii) Can we determine how much deviations from exclusive vertical transmission of traits will affect our ability to infer internal nodes?

Various approaches have been suggested to identify the role of horizontal transmission, some retaining and others abandoning tree-based approaches. For example, biologists explicitly model the possibility that not all genes within an individual will necessarily share the same branching phylogeny [reconciliation analysis, 20] [incomplete lineage sorting, 21], a method that can be applied to human cultural traits [19, 22, 23]. Cultural evolutionists turn to network analysis [24] or popularity spectra [25] to detect horizontal transmission, shown in each of these cited studies to be particularly predominant in oceanic environments. Other comparative social and evolutionary scientists turn to various matrix frameworks, employing multiple regression, Mantel tests and autologistic regression models to detect shared ancestry and/or cultural diffusion in their data [e.g., 26, 27-30]. While these latter alternatives can indicate the relative contributions of vertical versus horizontal transmission, there is generally no direct way to link them back to inform inferences of internal nodes.

Simulation studies are particularly helpful in examining the sensitivity of inferences about evolutionary processes to horizontal transmission. Nunn et al [31] studied character

evolution in a spatial framework, and showed that horizontal transmission can in some cases produce misleading inferences about evolutionary processes. Others show that identifying patterns of trait evolution, and trait values at internal nodes, depends not just of rates of borrowing but on whether traits are borrowed as packets or singly [32] [but see discussion in 33], or advise on focusing on less “unrealistic” rates of borrowing [34]. But, and this is the point, we rarely have direct windows onto the past, so speculations over whether traits are borrowed as packages or singly, and at what rate, are questionable [18, 19]. Decisions about what to consider a tolerable rate of borrowing or how to include borrowing explicitly in any analysis depends on what we know empirically, on the nature of the trait, and on whether the borrowing is global or local [as discussed in 24, 25, 31, 32, 33].

(b) Does the trait change over evolutionary or shorter time scales?

(i) Can we identify the rate at which societies change their cultural traits?

If cultural traits are highly facultative, and change over very short periods of time, they may not be amenable to phylogenetic reconstructions of their evolutionary history because there is simply too much variation at the tips of the trees. The signature of history might be weak compared to the strength of current selection pressures [35, 36]. Indeed, simulations show that high rates of evolutionary change have a strong depressive impact on measures of phylogenetic signal when determined through the fit of such traits onto an independently determined tree [the “true” tree, see 33]. Estimating highly facultative/flexible traits on an independently-derived tree risks obscuring the possibility that there were in fact multiple undetectable changes on each branch. A phylogenetic

approach cannot accurately reconstruct internal nodes if these do not retain a signal about the past [37] and might instead present a regression to the mean with larger changes earlier in the tree where deeper branches merge.

The problem of highly facultative traits is well-illustrated in a recent innovative study linking archaeological and contemporary ethnographic data to examine the association between dwelling size and post-marital residence [38]. By plotting dwelling size and residence rules onto a time-calibrated global super-tree of human populations the authors show that changes in house size precede changes in residence patterns. While there are possible explanations for these findings, it is hard to see intuitively how dwelling size (which might entail simply adding an extra room) is less mutable over evolutionary time than post-marital residence norms (which are known to vary with other traits such as forms of property transmission, lineality, and warfare). Dwelling size is by no means a uniquely facultative trait; polygyny, for example, also appears to be highly volatile [36, 39], raising the question of whether it is ever reasonable to put highly facultative traits on a phylogeny built on language evolution [35, 40, 41]?

Identifying rates of change and flexibility is complicated. Short of a time machine, direct evidence from history, archaeology or paleoanthropology is obviously the gold standard [41]. The preferred method for inferring past changes is to identify an independent source revealing the historical relationship among populations (step 1, as noted above), typically a linguistic or genetic tree with known branch lengths. Investigators then use the observed (tip) value of cultural traits to estimate (with probabilities) past values of the cultural trait

at its internal nodes, and where possible triangulating with independent sources of data, such as archaeological data [e.g., 42] or other well-known (and typically more recent) cultural sequences [such as the technological changes in brasswind cornets, 43]. However, for many behavioural and cultural traits such historical data are unavailable.

A different approach entails making predictions, both from intuition and empirical patterning, about the likely conservatism or volatility of traits in different domains: family and kin-based traits, for example, have long been held by anthropologists to change slowly [e.g., 44], or at least more slowly than traits directly linked to the environment [45, 46], but other patterns can emerge. Evidence from the Austronesian language family shows phylogenetic and geographic (or both) patterning of some of the social/kinship traits studied, but no clear model for just under half the 78 traits studied [11]. Similarly, material and technological traits can show widely varying phylogenetic signal [26, 46, 47].

(ii) Does the granularity at which cultural traits are defined affect the inferred rate of change?

Any measured rate of trait replacement depends on the granularity of trait measurement and categorization. This is particularly likely to be a problem for cultural traits, which are often reported as categorical states (such as bridewealth, dowry, brideservice, and no payments), states that are in many instances further combined into binary categories to retain statistical power for analysis [48]. Transition rates may reflect such granularity, for example a shift from matrilocality to non-matrilocality could take longer than a shift to multilocality, which encompasses matrilocality (see Section 2biii). The difficulty of inferring

transition rates is further compounded if the rate differs in different parts of the tree [49] or if it is higher for some transitions than for others [50]. Finally studies of contemporary cultural transitions, such as how bridewealth payments shift into effective dowry [51, 52] or how matrifocal institutions emerge and wither [53, 54], can throw light on the stability and/or changeability of these specific traits, but generalization to other contexts remains problematic.

In short, we still grapple with deriving *a priori* hypotheses for which cultural domains are most conserved as well as with developing sound methods for determining how rapidly they change. While the ancestral relationships among variants of cultural traits can be represented as trees using “the most parsimonious cladogram” [55, page 175], the calculation of indices to represent historical changes from such cladograms is problematic [33, Section 4d]. It is worth stressing that the vast majority of traits show little or no patterning with demographic and/or geographic indices that are supposed to reflect population histories [e.g., 11, 36, 56]. This all points to the importance of facultative adaptation across different domains, as envisaged by Rogers and Cashdan [35]. We should then, as previously emphasized [e.g., 19, 57], acknowledge huge variability in the rates at which traits change across domains, world regions, historical periods, and geographic scales before deciding whether or not a phylogenetic approach is appropriate for our particular trait of interest.

(iii) *Are the traits of interest measurable at a scale whereby the question can be satisfactorily answered?*

Closely related to the argument above, the grain at which traits are measured may preclude, or even distort, understanding of evolutionary dynamics [58]. Comparisons across large samples of diverse societies raise challenges for identifying traits that can be both measured in all societies and considered homologous. A trait that appears similar in different societies because it appears in the same context (e.g., ensuring food security in agriculturalists and foragers) might represent independent inventions or borrowing [18].

The granularity at which traits are measured is likely to affect our ability to detect sequences of cultural transitions, whether A is a necessary precondition to B or *vice versa*. For example, consider “social complexity”, a central trait in the debate over whether moral gods precede or postdate the emergence of complex societies [59, 60]. In an examination of the robusticity of the measure of social complexity Miranda and Freedman [61] show how two societies with radically different social organizations can look the same in terms of a unidimensional measure of social complexity. By exploring beyond the first principal component of a society’s social complexity ranking [see also 62], the authors develop a scale that could add considerable nuance to the conclusion (made from transition analyses) that societal complexity causally preceded moralizing gods. The fact that often only coarse codings are available in no way invalidates phylogenetic (or any other comparative) analyses, but clearly impedes the strength of precise inference, particularly with respect to sequences [48]. Furthermore, it bears stressing that breaking down traits into multiple states will likely make it harder to identify phylogenetic signal, if finer grain traits are more likely to change (as discussed in Section 2bii).

More generally, for each of these data-related problems discussed in this section, starting from a clear theoretical framework that outlines the specific predicted relationships among well-defined variables can help to decide whether the empirical data available are adequate to the question at hand. For example, to understand the evolution of a concept such as “social complexity,” now that social scientists recognize its general patterning we might learn more from examining the distribution of specific social traits associated with complexity [as in 63, for non-humans] rather than continue to employ coarse unidimensional variables. Where binarization is unavoidable, it should be based on clear theoretical justification.

(c) Are there appropriate independently-derived population trees available for phylogenetic analysis?

(i) Is there an appropriate tree to study the trait of interest at the scale of interest?

Assuming use of a phylogenetic tree is deemed appropriate, the next question is whether the tree available well reflects the demographic history of the culture-bearing population. With the use of genetic or, more commonly, linguistic trees, an affirmative answer seems reasonable insofar as the associations between populations and their languages are generally quite stable. However, such detailed trees can generally only be constructed for a small number of closely related populations. Super-trees [64], which patch together multiple of these trees into a global phylogeny [65], offer the potential to explore the generality of explanatory hypotheses for cultural diversity, but they may incorporate very different histories in different parts of the tree. These differences could reflect the distinct data sources from each branch of the tree, or genuine lineage-specific evolutionary

patterns, to which we turn below. It is also important to consider whether the trees chosen to represent population histories reflect a time scale appropriate to the social learning and transmission processes that generated the data being evaluated by the comparative method.

(ii) Are there lineage-specific or adaptive effects where trait changes differ in different parts of the sample?

Traits may show different evolutionary patterning in different language families and/or in different parts of the world. Take, for example, post-marital residence (whether a new married couple live with his, her or neither set of parents). Given the range of ecological, technological, and institutional factors that have been shown to influence this cultural pattern, it is hardly surprising that there are contradictory findings between different studies conducted on different samples [as discussed in 50, 66]. These can usefully be thought of as lineage-specific effects with respect to the patterning of evolutionary transitions [49]. As long as the sample sizes in each part of the sample are sufficient for inference, this diversity should be seen as a strength of phylogenetically-based global comparisons, providing material with which to sharpen our understanding of the precise conditions under which particular evolutionary transitions occur. In short while there may be a justification for aiming to test classical hypotheses at increasingly global scales [e.g., 60] failure to find generalized support may actually be highly informative. In similar vein we might expect to find different evolutionary dynamics in agricultural versus foraging societies, rendering problematic inferences made from current samples that are typically

dominated by agriculturalists, who (as in our opening metaphor) offer only a limited palette.

In cases where we know or suspect the specific factors that influence cultural patterns differently across the tree, we can be more explicit about evolutionary processes. For example, Ross et al. [67] first built an explicit evolutionary model to assess how the classification of societies as stratified or not might influence the origin and maintenance of the practice of female genital modification. Next, they used an approach based on the Ornstein-Uhlenbeck process [68] which allows for adaptive hypotheses to be evaluated with phylogenetic models that explicitly include adaptive dynamics, both to test for coevolution across the phylogeny and to contrast the strength of selection on a trait with that of drift and other selective processes. In addition to testing explicit co-evolutionary hypotheses, external knowledge about the influence of a factor on the cultural trait can also help to reduce the error in reconstruction of the history of the trait.

(iii) Are inferences affected by the resolution and structure of the tree?

Phylogenetic relationships among an increasing number of societies are starting to be described, primarily based on language patterns [6], and increasingly incorporated into supertrees [e.g., 69]. Details of the phylogeny are important to understanding how much information about evolutionary processes can be extracted [70]. One issue is uncertainty in the tree: researchers need to consider carefully whether their tree provides a clear branching pattern, as well as robust dates for when the splits occurred. Analyses can now incorporate uncertainty, but often at a reduction in the power to make inferences [71]. In

addition, the shape and the size of the phylogeny influence the inferences that can be made. Small sets of societies, which branch off very early, can become very influential for the reconstruction of internal nodes, the same way that outgroups are used to reconstruct the state of a common ancestor of a clade [72], if data are available for only few societies in the main clade. More generally, balanced trees, with regular branching patterns across the tree, appear more robust to phylogenetic inferences of cultural traits, especially in the presence of horizontal transmission, than unbalanced trees [73]. In contrast, phylogenies with early bursts of speciation, with long branches leading to the societies, contain less information to extract the correct patterns of changes in cultural traits [70].

(d) Is this the right sample from which to make inferences?

(i) Is the sample large enough?

Finding societies, or ethnolinguistic units, with appropriate measures for a cultural evolutionary analysis can be challenging. The vast majority of research to date is based on language families with well-resolved trees, although the number of societies available with appropriate data is often quite small. So, for example, in a recent study of post-marital residence Indo-European provided 27 societies, Uto-Aztecan 26 societies, Pama-Nyungan 66 societies, Bantu 69 societies, and Austronesian 135 societies [49]. In a different kind of study investigating the association between parasite stress and political traits like authoritarianism and democracy data were available only from 52 nation states [30]. The issue is not just the number of societies with data (tips), but the total number of societies in a clade, and the patterning of variation among them: if variant A is present in

all societies in one clade and trait B present in all societies of the other clade, most likely only one change occurred, making it impossible to associate any factor with that change statistically. Indeed, Moravec et al [49] had to drop the entire Indo-European family from their study because there were no matrilineal populations. The obverse (too many changes) is also a problem: if each society differs from its neighbour, it is impossible to reconstruct changes in the past — in our metaphor, the picture of the past is murky brown. In summary, available sample sizes may not allow for reliable inferences.

(ii) What about missing data and extinction?

Researchers might have to drop cases from their phylogeny, or sample societies strategically, on the basis of trait data availability. Missing data are an issue for at least two reasons. First, if information is available for many societies in one clade but only a few societies in other clades, trees will be highly unbalanced (see Section 2ciii). Second, data (or branches) may be missing not at random with respect to the trait of interest. With respect to the first issue, in Hrnčír et al's study [38] the sample for studying the association between dwelling size and post-marital residence was heavily biased to New World sites, on account of the scant attention paid to material traits (such as house size) by European as opposed to American ethnographers. While the resulting sample may not be biased with respect to the hypothesis (i.e., data are missing at random) the effect of such uneven and sparse sampling might obscure any phylogenetic signal in the variables and influence inferences about internal nodes. With regard to the second problem, where there is an absence of records (e.g., among foragers in highly productive habitats, see

below), inferences must be drawn with great caution because, following our metaphor, we simply do not have the full palette of colours.

Extinctions, including unknown ones, are a linked concern insofar as simulation studies show that high extinction rates bring considerable error to estimates of root states [45], as well as influence Type I and II errors [31-33]. Furthermore, extinctions are unlikely to be random. The reason a society is extinct (i.e., does not appear at a tip) may well be associated with its trait values prior to extinction. For example, examining prevalent marriage customs (e.g., are they arranged, are they polygynous) on a tree of contemporary hunter-gatherers [74] may bias ancestral estimates toward the traits of populations that have managed to persist, often in highly marginal areas, into the present era. Suroweic et al [50] reconstructed, both in their global super-tree and in the Bantu language family, the ancestral state for lineality to be non-matrilineal and the ancestral state for post-marital residence to be non-matrilocal [in the case of Bantu contrary to inferences from historical and linguistic data, 75] Could this be because matrilineal (and matrilocal) societies are less likely to have survived shifts to intensive agriculture and pastoralism, which typically favour male-centred kinship institutions?

3. Assessing potential challenges through simulation

(a) Simulations to reveal both general and specific limitations

Various studies have used simulations of the likely processes that generate and modify traits to assess the challenges discussed above. These typically focus on the general feasibility of cultural macroevolution studies [e.g., 31, 45] or the limitation of particular

approaches to recover the history of traits [e.g., 70, 76, 77]. Here, following previous suggestions [e.g., 16, 34], we show that simulations can be usefully employed to assess the extent to which these challenges might influence a particular study.

Our simulations investigate the potential impact that the different challenges might have on phylogenetic reconstructions of trait evolution. We focus on three types of inferences that are common in studies of the phylogenetic reconstruction of trait evolution (Figure 1): are some changes between the states of the trait favoured (inference i)?; do socioecological conditions influence which changes among the states of the trait occur (inference ii)?; are there differences between lineages in which changes among the states of the trait occur (inference iii)? In trying to address these inferences, there is a risk of concluding that the answer to the inference is yes even though in truth there was no effect (false positive Type I error) and of concluding that the answer to the inference is no even though there was an effect (false negative Type II error). Our simulations estimate the expected chance of obtaining false positive or false negative conclusions given a particular phylogeny and dataset. In addition, following from the discussion above, we examine how these error rates are influenced by properties of the sample (sample size, missing data), properties of the tree (tree shape), and properties of the traits (rate of change, number of variants, horizontal transmission) (Figure 1). By simulating likely histories of the trait of interest across the given phylogeny, the power to discern alternative histories can be assessed.

We demonstrate this approach using two language phylogenies, one for societies in the Western North American Indian (WNAI) dataset [78] (Figure S1) and one for societies included in a recent phylogenetic study of the Pama-Nyungan language family [79] (Figure S2). For the WNAI, we constructed a phylogenetic tree for the 172 societies based on a hierarchical language classification [80, modified by 81]. For analyses that require bifurcating trees, we resolved polytomies in the phylogenies randomly, but did not represent this uncertainty by repeating the simulations across multiple resolutions because our aim here is to illustrate example cases. For the 306 Pama-Nyungan societies, we used the phylogeny provided in the study. The results we present are dataset-specific, so inferences cannot be directly transferred to other studies. However, several of the findings are consistent across the two phylogenies, and we emphasize we selected language trees that largely reflect the sample size, tree shape and trait distributions that cultural evolutionists typically encounter. We provide all input data, code and output (https://github.com/dieterlukas/CulturalMacroevolution_Simulation). The code can be adapted for use on any phylogeny prior to analysis to assess the specific power of a given dataset to test phylogenetic hypotheses.

433 **Figure 1: Common inferences in cultural macroevolution and their associated errors**

434 We focus on three inferences about the evolution of cultural traits, and the errors that can occur when trying to address

435 them.

<u>Inference</u>	<u>i) direction of change</u>	<u>ii) socioecological influences</u>	<u>iii) lineage differences</u>
Question	Are all changes equally likely or are some changes favoured?	Do socioecological conditions influence which changes occur?	Are there differences between lineages in which changes occur?
Examples	<ul style="list-style-type: none"> - Does social complexity increase in a series of similar sequential steps? - Can matrilineality change directly to patrilineality or does it have to pass through ambilineality? 	<ul style="list-style-type: none"> - Does the presence of social stratification influence whether female genital modification practices occur? - Does subsistence mode influence sex biases in residence patterns? 	<ul style="list-style-type: none"> - Does the pathway from patrilineality to matrilineality differ in different lineages? - Are changes in grammatical features universal or language-dependent?
Type I error (false positive)	<p><u>Actual history:</u> the trait changed randomly between all states of the trait</p> <p><u>Wrongly inferred history:</u> changes between certain states happened more or less frequently than other changes</p>	<p><u>Actual history:</u> changes in the trait did not depend on the ecological conditions</p> <p><u>Wrongly inferred history:</u> ecological conditions influenced changes between certain states to occur more frequently</p>	<p><u>Actual history:</u> changes in the trait did not differ between different parts of the phylogenetic tree</p> <p><u>Wrongly inferred history:</u> changes between certain states happened more or less frequently than other changes</p>
Type II error (false negative)	<p><u>Actual history:</u> selection favoured changes between certain states over other changes</p> <p><u>Wrongly inferred history:</u> all changes occurred with equal likelihood</p>	<p><u>Actual history:</u> Changes to certain states were more common in some ecological conditions than others</p> <p><u>Wrongly inferred history:</u> changes did not depend on ecological conditions</p>	<p><u>Actual history:</u> Changes to certain states are more common in some parts of the tree</p> <p><u>Wrongly inferred history:</u> changes occurred equally in all parts of the tree</p>

(b) Methods for our simulations of cultural macroevolution

(i) Simulating discrete and continuous traits

Across the two phylogenies, we simulated the evolution of discrete (having either two or four different states) and continuous traits using functions of the package ‘phytools’ in R [82]. Simulations start at the root (with the first discrete variant or the value zero for continuous traits), and estimate transitions along all branches. Transitions between different states (discrete traits) or rates of change (continuous traits) occurred according to four different scenarios. In the first scenario (to obtain the rate of Type I error rates of wrongly inferring directional changes, of wrongly inferring socioecological influences on transitions, and of wrongly inferring lineage differences), all transitions and changes were the same and occurred randomly and with one consistent probability, reflecting a null model drift scenario, where all changes are equally likely across the whole tree. In the second scenario (to obtain the Type II error rate of wrongly missing directional change), there are two rates of transitions and changes. For discrete traits, one rate reflects higher transitions towards one of the variants from the other one or three variants whereas the other rate reflects all other transitions. For continuous traits, increases are more likely than decreases. In the third scenario (to obtain the Type II error rate of wrongly missing socioecological differences), transition rates differ between societies associated with different socioecologies, such that one variant (discrete traits) or positive values (continuous traits) are favoured in branches within lineages associated with different socioecologies. To represent different socioecological conditions as potential correlates of the changes in the simulated cultural trait, we classified the WNAI societies as either living in forest habitats or not, and the Pama-Nyungan societies as hunter-gatherers or

food-producers. In the fourth scenario, transition rates differ between lineages, with changes towards one of the variants (discrete traits) or increases towards positive values (continuous traits) more common along the branches of the largest clade compared to the rest of the tree (to obtain the Type II error rate of wrongly missing lineage differences).

For all scenarios, we extract the simulated states of the trait across the tips of the phylogeny. The reported error rates reflect the number of independent simulations in which the analysis supported the wrong inference. The error rates for the baseline (Figure 2A) are based on 110 simulations each, reflecting the different settings and ten independent repetitions of each setting. The error rates for simulations assessing the influence of properties of the sample, of the tree, and of the trait are based on the respective subsets of simulations, and we report how these deviations from our chosen baseline change the proportion of wrong inferences (indicated by + and – designations in Figure 2B-D and Supplementary Figures S3-S5).

Figure 2: Sources and rates of errors across the simulations of trait evolution for the WNAI and the Pama-Nyungan societies

The illustrations show examples of potential sources of errors in phylogenetic reconstructions of trait evolution. They depict the evolutionary history of a trait with two states (red and black, representing either a discrete trait or negative and positive values of a continuous trait), the resulting values at the tip, and most likely reconstruction. The values underneath show the error rates in baseline simulations (A), and how much these change (increase + or decrease -) depending on an example property of the sample (missing data, B), an example property of the trait (random horizontal transmission, C), and an example property of the tree (a late-burst phylogeny, D). The reported error rates display the percentage of independent repetitions in which a phylogenetic reconstruction of trait evolution either wrongly supported an evolutionary model in which rates of change differed across the tree even though all changes were simulated to occur randomly (Type I error, false positive); or wrongly did not support an evolutionary model in which rates of change differed even though this was how the change was simulated (Type II error, false negative).

A) Baseline simulation

tree is undated for WNAI, dated for Pama-Nyungan

With a standard approach, inferences are probabilistic, and might occasionally lead to errors deeper in the tree

error rates discrete trait

direction of change	socioecological influences		lineage differences	
Type I: false positive	Type I: false positive		Type I: false positive	
WNAI 12% PamaNy 15%	WNAI 40% PamaNy 29%		WNAI 38% PamaNy 26%	
Type II: false negative	Type II: false negative		Type II: false negative	
WNAI 52% PamaNy 36%	WNAI 65% PamaNy 46%		WNAI 54% PamaNy 60%	

error rates continuous trait

direction of change	socioecological influences		lineage differences	
Type I: false positive	Type I: false positive		Type I: false positive	
WNAI 13% PamaNy 0%	WNAI 15% PamaNy 51%		WNAI 6% PamaNy 48%	
Type II: false negative	Type II: false negative		Type II: false negative	
WNAI 42% PamaNy 51%	WNAI 1% PamaNy 2%		WNAI 0% PamaNy 0%	

B) Example of property of the sample

data is missing for one variant

If there are biases in which data are available, it might lead to wrong inferences of internal nodes and reduces power to detect associations

error rates discrete trait

direction of change	socioecological influences		lineage differences	
Type I: false positive	Type I: false positive		Type I: false positive	
WNAI +1% PamaNy +/-0%	WNAI -2% PamaNy +18%		WNAI +1% PamaNy +2%	
Type II: false negative	Type II: false negative		Type II: false negative	
WNAI -1% PamaNy +2%	WNAI +8% PamaNy +46%		WNAI -3% PamaNy +1%	

error rates continuous trait

direction of change	socioecological influences		lineage differences	
Type I: false positive	Type I: false positive		Type I: false positive	
WNAI +20% PamaNy +/-0%	WNAI +3% PamaNy +7%		WNAI -6% PamaNy -33%	
Type II: false negative	Type II: false negative		Type II: false negative	
WNAI +2% PamaNy +31%	WNAI +10% PamaNy +8%		WNAI +/-0% PamaNy +/-0%	

C) Example of property of the trait

random horizontal transmission

A high rate of independent horizontal transmission will introduce outliers into clades: for rare binary traits this can lead to spurious associations while for continuous traits it will obscure patterns

error rates discrete trait

direction of change	socioecological influences		lineage differences	
Type I: false positive	Type I: false positive		Type I: false positive	
WNAI +39% PamaNy +31%	WNAI +1% PamaNy +18%		WNAI +1% PamaNy +15%	
Type II: false negative	Type II: false negative		Type II: false negative	
WNAI -21% PamaNy -13%	WNAI -3% PamaNy -1%		WNAI -20% PamaNy -18%	

error rates continuous trait

direction of change	socioecological influences		lineage differences	
Type I: false positive	Type I: false positive		Type I: false positive	
WNAI +77% PamaNy +78%	WNAI -8% PamaNy -48%		WNAI +28% PamaNy +21%	
Type II: false negative	Type II: false negative		Type II: false negative	
WNAI -30% PamaNy -26%	WNAI +94% PamaNy +95%		WNAI +/-0% PamaNy +/-0%	

D) Example of property of the tree

late-burst phylogeny with divergent lineages

Lineages will show differences by chance if they are separated by long branches, and if variation in ecology arose in all lineages short branches do not provide sufficient time for selection to lead to a response

error rates discrete trait

direction of change	socioecological influences		lineage differences	
Type I: false positive	Type I: false positive		Type I: false positive	
WNAI -4% PamaNy -3%	WNAI -21% PamaNy +43%		WNAI +18% PamaNy +7%	
Type II: false negative	Type II: false negative		Type II: false negative	
WNAI +13% PamaNy +15%	WNAI +28% PamaNy +52%		WNAI -8% PamaNy +11%	

error rates continuous trait

direction of change	socioecological influences		lineage differences	
Type I: false positive	Type I: false positive		Type I: false positive	
WNAI +46% PamaNy +62%	WNAI +/-0% PamaNy -2%		WNAI +34% PamaNy +13%	
Type II: false negative	Type II: false negative		Type II: false negative	
WNAI -6% PamaNy -23%	WNAI +90% PamaNy +60%		WNAI +/-0% PamaNy +/-0%	

(ii) *Analysing simulated traits*

To determine whether some changes in the simulated *discrete* traits occurred more frequently than other changes (inference i), we compare the likelihood of a reconstruction in which all transitions are assumed to occur at the same rate to that of a reconstruction in which some transitions occur more frequently than others. We estimate the likelihoods of the equal and the biased models using the function ‘ace’ of the package ‘ape’ [83] in R, assessing significance with a likelihood ratio test adjusting the degrees of freedom to reflect the different number of parameters in the two models. To determine whether some changes occurred more in specific socioecological conditions (inference ii), we compare the likelihood of a reconstruction where transitions in the simulated trait are independent from changes in the socioecological condition to a reconstruction where changes in the simulated trait depend on the socioecological condition. We used the function ‘discrete’ in the software Bayestraits [84] to estimate the likelihood of a dependent model and the likelihood of an independent model, comparing them using a likelihood ratio test. To determine whether changes were more likely to have occurred in the largest clade compared to the rest of the tree (inference iii), we estimate the likelihood of a reconstruction where transitions in the simulated trait are independent of where they occur in the tree and the likelihood of a reconstruction where transitions in the simulated trait are dependent on whether they occur on a branch in the largest clade or on a branch in a different part of the tree. For this we also used the function ‘discrete’ in the software Bayestraits and conducted a likelihood ratio test.

To determine whether changes in the simulated *continuous* traits occurred predominantly in one direction rather than randomly (inference i), we compared the likelihood of a reconstruction in which changes occurred as under a null Brownian drift model to a reconstruction in which changes occurred with a trend in one direction using functions of the package 'OUwie' [85] in R, assessing significance with a likelihood ratio test. To determine whether values differed between societies in different socioecological conditions (inference ii), we performed phylogenetic regressions to assess whether socioecological conditions are associated with differences in the simulated values observed at the tips [using functions of the package 'MCMCglmm' in R, 86], assessing significance using the p-value). To determine whether values in the largest clade differed from the values in the remainder of the tree, we performed phylogenetic regressions to compare all values in the largest clade to the remaining values (using functions of the package 'MCMCglmm' in R, and assessing significance using the p-value).

(iii) Identifying the effect of differences in the properties of the trait, of the sample, and of the tree

In addition to a baseline scenario, we varied the following parameters during the simulations: properties of the trait, of the sample, and of the tree. To change *properties of the trait* we: a) set the number of possible variants of the trait to be either two or four (discrete traits), b) changed the rate at which transitions could occur from low to high (for a slow rate, this equals to an average of 10 changes across the 610 branches in the Pama-Nyungan phylogeny and about half as many changes across the 342 branches of the WNAI phylogeny; for a fast rate, this equals to an average of 75 changes across the

Pama-Nyungan phylogeny and about 40 changes across the WNAI phylogeny), and c) introduced horizontal transmission by having 10% societies at the tips copy a variant of another random society, another society in the same ecology, or their closest geographic neighbour. To change *properties of the sample*, we investigated samples with: a) the full set of societies, b) only societies in the largest clade, and c) 25% or 50% of data missing from a random set of societies, from societies in the dominant ecology, or from societies with one of the variants (discrete traits) or with positive trait values (continuous traits). To change *properties of the tree*, we manipulated branch lengths: a) to reflect trait change at the time of split, where branches lengths are all the same, b) to reflect an early burst into the existing societies, where branches between clades are short and branches to the tips are long, and c) to reflect a late radiation where branches separating clades are long and all diversification within the clades happened very recently. We display examples of manipulations of properties of the trait, of the sample, and of the tree in Figure 2 and below; for additional results see Supplementary Materials.

(c) Findings from our simulations of cultural macroevolution

Our simulations show that for these particular datasets and phylogenies, properties of the trait have the biggest influence on the frequency of erroneous inferences, properties of the sample have a relatively minor influence, while the influence of properties of the tree are context-specific (Figure 2).

We set up the baseline simulations to generate data similar to those commonly available in cultural phylogenetics, and we analyzed the simulated data with the standard settings of the respective methods. In these baselines, we obtained substantial false positive (Type I) and false negative (Type II) errors that range between 12%-40% and 36%-65% respectively for discrete traits, and 0%-51% and 0%-51% for continuous traits (Figure 2A). Continuous traits generally show lower false negative errors than discrete traits: the amount of variation in discrete traits is restricted, reducing power to detect weak effects, in particular when relying on the standard settings of the analyses which attempt to estimate a large number of parameters without any prior information. False positive rates are largely comparable for continuous and discrete traits: the simulated traits changed relatively rapidly, such that by chance lineages and ecologies will differ. There is no consistent difference in performance between the two language trees, despite their difference in shape and tree structure balance. In the following, we present changes in the errors contingent on various properties of traits, trees and sample, but we should emphasize that throughout errors are high. While undoubtedly some of these problems might be specific to the situations we chose to simulate, and the precise magnitude of the errors might be reduced by adapting the settings of the analyses to independently available information (such as likely types of transmission), the size of these errors should not be overlooked.

Regarding directional change, the false positive error of wrong inference was increased most markedly from the baseline rate when variants of the trait were transmitted horizontally among societies (Figure 2C, Figure S3). Such horizontal

transmission might reflect actual directional changes, but we found relatively high error rates even when horizontal transmission occurred randomly: relatively high rates of random horizontal transmission (in our simulations 10% of extant societies copy a variant) will introduce highly divergent variants, leading to Type I errors of wrongly supporting a positive signal as these divergent values would require faster changes than assumed under a null model of random drift. The Type II error of wrongly missing directional change was generally high both for discrete and continuous traits: it was highest when power was reduced because there are very few changes in total across the phylogeny, that is with fewer variants of the trait (Figure S5) and lower when horizontal transmission increased the frequency of certain states throughout the tree (Figure 2C) .

Regarding the detection of socioecological influences on change, although the false positive error rate was already high in the baseline for both discrete and continuous traits, it was larger when there was horizontal transmission that could increase the occurrence of a rare variant within one ecology (Figure S4) and when the societies in different ecologies are separated by long branches such that changes might occur by chance (e.g. late burst phylogeny for the Pama-Nyungan, Figure 2D). The false negative error rate of missing socioecological influences on changes was highest when data was missing non-randomly (Figure 2B), in particular if certain variants are rare to begin with (less than 10% of Pama-Nyungan societies are food-producers, Figure S2). False negative error rates for detecting a socioecological influence also increased under random horizontal transmission (Figure 2C) and late-burst phylogenies (Figure 2D) in both language families, although only for continuous traits.

Regarding lineage-specific patterns, the false positive error rate of wrongly inferring lineage differences on changes was already high in the baseline simulation for both discrete and continuous traits. It was considerably increased when there was structured horizontal transmission within lineages or ecologies (Figure S3), random horizontal transmission (Figure 2C), and as a result of tree structure, specifically in late-burst phylogenies where long branches separate lineages such that differences might occur by chance (Figure 2D). The false negative error rate of wrongly missing lineages differences was already high in the baseline (for discrete traits) and appears to decline when horizontal transmission occurs (Figure 2C, Figure S3).

4. Discussion and Outlook

The most striking aspect of these simulations is the very high rate of false positive Type I and false negative Type II errors contingent on properties of the traits, of the tree shape, and less so of the sample. While this result may be specific to the WNAI dataset (with its poorly resolved language phylogeny, its categorical codings, and its unbalanced clade structures), largely similar false positive and false negative errors in the Pama-Nyungan analyses suggest they might be quite general.

The strong effects of horizontal transmission on false inferences about selection may reflect the fact that in the WNAI linguistic and spatial signals are relatively uncorrelated [26, see also 39]. This is not always the case; phylogenies vary in the extent to which they can reliably capture the spread of cultural traits, which will depend on whether or not there

has been a well-documented linguistic and cultural expansion as in the Austronesian and Bantu language families. While we might all hope that the WNAI-specific findings are not generalizable, the analysis of Pama-Nyungan (with its better-balanced clade structure) is concerning in that it shows very similar false inferences resulting from horizontal transmission.

Despite the fact that cultural phylogeneticists differentiate, quite rightly, the distinct objectives of different cultural phylogenetic studies, our simulation results reveal problems that are not particular to any one objective. More specifically, we show that properties of the traits, tree and (less so) sample generate inferential problems that result from inaccurate reconstructions of internal nodes. These inaccuracies impact *not just studies that seek determination of ancestral states*, but also those concerned with the identification of independent adaptations, the understanding of potential sequences of cultural transitions, and the determination of whether transitions are correlated with changes in other traits, *insofar as each of these objectives requires estimation of internal nodes*.

Given the potential generality of these effects of properties of traits (and to a lesser extent of the tree and the sample) on the validity of inferences, and the relevance of these findings for the multiple objectives within cultural phylogenetic research, we suggest that the concerns we review in Section 2 continue to be legitimate. Accordingly, we offer a simulation that identifies the expected error rates associated with a large set of trait, tree and sample properties (see code on GitHub) which we encourage investigators to use to

determine the error rates they might expect in their particular analysis. We view our contribution here as complementary to that of staunch supporters of phylogenetic analyses who have themselves acknowledged that, where cultural reticulation has been rampant, projecting traits onto trees can only produce “meaningless results” [34, page 366], adding that phylogenetic analysis may not be suitable for all temporal and spatial time scales [3].

We also propose the following items as a starting point for a checklist when planning a cultural macroevolution study. (1) Research questions should be informed by solid theoretical models [e.g., 67, 87]. This will help address many of the dilemmas and tradeoffs outlined here – specifically with respect to how to identify which sample to use, and how to measure the trait of interest. (2) Investigators should consider carefully whether the tree chosen to represent population histories reflects a time scale appropriate to the social learning and transmission processes that generated the data being evaluated by the comparative method. Ideally, the tree will reflect diverse sources of data across history, linguistics, genetics and archaeology [see also 41]. (3) Investigators should consider carefully how the trait of interest is transmitted, using ethnographic or historical materials where available, taking care to differentiate mechanisms of transmission for between and within populations; this will inform about the potential rate of change, and whether horizontal transmission is likely to be prevalent [see also 88]. (4) Investigators should explore how missing data, horizontal transmission, and/or differential lineage extinction might influence their inferences for their particular sample (see our code). Investigators may have additional information about the transmission of the cultural trait

674 they are studying that can inform the analyses and there are multiple ways to model such
675 evolutionary processes. (5) Investigators should acknowledge the centrality of internal
676 node trait estimation to their inferences. For example, researchers often assert that if they
677 are using phylogenies only to control for statistical independence the precise details of
678 internal node reconstructions are not critical. This may be true, but it is incumbent then
679 not to draw inferences of causality from their phylogenetic analyses when, as we have
680 shown, certain features of the tree, the sample or the traits may exacerbate these
681 inaccuracies.

682

683 To return to our metaphor, we urge investigators to consider carefully what the picture
684 they are painting of the past is actually going to show. While it may not offer perfect
685 resolution, nor the full colour spectrum, is there sufficient detail to discern the key features
686 of interest?

687

688 In sum, we are not arguing against phylogenetic studies of cultural traits. Rather, we hope
689 our overview and our simulation will encourage researchers to consider persistent perils
690 inherent in answering questions about the history of a cultural trait. While we recognize
691 the importance of the cultural evolutionary goal of putting “pre-history back into
692 anthropology” [11, page 406], we suspect more caution is still needed. Evidence for our
693 position lies in the increasingly contradictory findings that are emerging in the literature
694 regarding ancestral states, sequence of changes, and adaptive value. The very fact that
695 the field has grown enough to produce discrepant findings is positive and generates
696 progress. Discrepancies may well reflect lineage, region or scale-specific effects, as well

as differences in the definition or classification of traits which, with further research, will only increase our understanding of the drivers of cultural diversity, but we do need to keep in mind that methodological problems might also be entailed.

Acknowledgments

We thank the members of the Department of Human Behavior, Ecology & Culture at the MPI EVA for feedback during the planning of this study, and Fiona Jordan, Sam Passmore and Tom Currie for helpful comments on an earlier version of this manuscript. We also thank the organizers of this special issue, particularly Eva Boon, as well as the Lorentz Center at Leiden University for hosting the Foundations of Cultural Evolution workshop.

References

- [1] Pagel, M. & Mace, R. 2004 The cultural wealth of nations. *Nature* **428**, 275-278.
- [2] Borgerhoff Mulder, M., Nunn, C. L. & Towner, M. C. 2006 Macroevolutionary studies of cultural trait transmission. *Evolutionary Anthropology* **15**, 52-64.
- [3] Gray, R. D. & Watts, J. 2017 Cultural macroevolution matters. *Proceedings of the National Academy of Sciences* **114**, 7846-7852.
- [4] Lipo, C. P., O'Brien, M. J., Shennan, S. & Collard, M. 2006 Mapping Human History: Culture, Language and Biology. (New Jersey, Aldine Transaction.
- [5] Mace, R., Holden, C. J. & Shennan, S. 2005 The Evolution of Cultural Diversity: A phylogenetic Approach. (London, UCL Press.
- [6] Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H.-J., Blasi, D. E., Botero, C. A., Bowern, C., Ember, C. R., et al. 2016 D-PLACE: A Global Database of Cultural, Linguistic and Environmental Diversity. *PLOS ONE* **11**, e0158391. (DOI:10.1371/journal.pone.0158391).
- [7] Boas, F. 1940 *Race, Language and Culture*. New York, The MacMillan Company.
- [8] Murdock, G. P. 1949 *Social structure*. New York, Free Press.

- [9] Moylan, J. W., Borgerhoff Mulder, M., Graham, C. M., Nunn, C. L. & Hakansson, T. 2006 Cultural traits and linguistic trees: Linguistic signal in East Africa. In *Mapping Human History: Phylogenetic Approaches in Anthropology and Prehistory* (eds. C. P. Lipo, M. J. O'Brien, S. Shennan & M. Collard), pp. 33-52. New York, Aldine Transaction.
- [10] Jordan, P. & Shennan, S. J. 2003 Cultural transmission, language, and basketry traditions amongst the California Indians. *Journal of Anthropological Archaeology* **22**, 43-74.
- [11] Mace, R. & Jordan, F. M. 2011 Macro-evolutionary studies of cultural diversity: a review of empirical studies of cultural transmission and cultural adaptation. *Philosophical Transactions of the Royal Society B* **366**, 402-411.
- [12] Bortolini, E., Pagani, L., Crema, E. R., Sarno, S., Barbieri, C., Boattini, A., Sazzini, M., da Silva, S. G., Martini, G., Metspalu, M., et al. 2017 Inferring patterns of folktale diffusion using genomic data. *Proceedings of the National Academy of Sciences* **114**, 9140-9145. (DOI:10.1073/pnas.1614395114).
- [13] Spolaore, E. & Wacziarg, R. 2009 The diffusion of development. *Quarterly Journal of Economics*, 469-529.
- [14] Kamilar, J. M. & Cooper, N. 2013 Phylogenetic signal in primate behaviour, ecology and life history. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**, 20120341.
- [15] Blomberg, S. P., Garland, T. & Ives, A. R. 2003 Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* **57**, 717-745.
- [16] McElreath, R. 1997 Iterated parsimony: A method for reconstructing cultural histories [M. Sc. thesis], Department of Anthropology, University of California at Los Angeles.
- [17] Lycett, S. J. 2015 Differing patterns of material culture intergroup variation on the high plains: quantitative analyses of parfleche characteristics vs. moccasin decoration. *American Antiquity*, 714-731.
- [18] Boyd, R., Borgerhoff Mulder, M., Durham, W. H. & Richerson, P. J. 1997 Are cultural phylogenies possible? In *Human by Nature: Between Biology and the Social Sciences* (eds. P. Weingart, S. D. Mitchell, P. J. Richerson & S. Maasen), pp. 355-386. Mahwah, NJ, Erlbaum.
- [19] Matthews, L. J., Tehrani, J. J., Jordan, F. M., Collard, M. & Nunn, C. L. 2011 Testing for Divergent Transmission Histories among Cultural Characters: A Study Using Bayesian Phylogenetic Methods and Iranian Tribal Textile Data. *PLoS ONE*. (DOI:doi:10.1371/journal.pone.0014810).
- [20] Page, R. & Charleston, M. 1998 Trees within trees: Phylogeny and historical associations. *Trends in Ecology and Evolution* **13**, 356-359.
- [21] Maddison, W. P. & Knowles, L. L. 2006 Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* **51**, 21-30.
- [22] Verkerk, A. 2019 Detecting non-tree-like signal using multiple tree topologies. *Journal of Historical Linguistics* **p**, 9-69.
- [23] Tehrani, J. J., Collard, M. & Shennan, S. 2010 The cophylogeny of populations and cultures: reconstructing the evolution of Iranian tribal craft traditions using trees and jungles. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365**, 3865-3874.
- [24] Gray, R. D., Bryant, D. & Greenhill, S. J. 2010 On the shape and fabric of human history. *Phil. Trans. R. Soc. B* **365**, 3923-3933. (DOI:doi:10.1098/rstb.2010.0162).
- [25] Nakamura, M., Wakano, J. Y., Aoki, K. & Kobayashi, Y. 2020 The popularity spectrum applied to a cross-cultural question. *Theoretical population biology* **133**, 104-116.
- [26] Towner, M. C., Grote, M. N., Venti, J. & Borgerhoff Mulder, M. 2012 Cultural Macroevolution on Neighbor Graphs: Vertical and Horizontal Transmission among Western North

- American Indian Societies. *Human Nature* **23**, 283-305. (DOI:10.1007/s12110-012-9142-z).
- [27] Gavin, M. C., Kavanagh, P. H., Haynie, H. J., Bown, C., Ember, C. R., Gray, R. D., Jordan, F. M., Kirby, K. R., Kushnick, G. & Low, B. S. 2018 The global geography of human subsistence. *Royal Society open science* **5**, 171897.
- [28] Beheim, B. A. & Bell, A. V. 2011 Inheritance, ecology and the evolution of the canoes of east Oceania. *Proceedings of the Royal Society B: Biological Sciences* **278**, 3089-3095.
- [29] von Cramon-Taubadel, N. & Lycett, S. J. 2018 Assessing the relative impact of historical divergence and inter-group transmission on cultural patterns: a method from evolutionary ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20170054.
- [30] Bromham, L., Hua, X., Cardillo, M., Schneemann, H. & Greenhill, S. J. 2018 Parasites and politics: why cross-cultural studies must control for relatedness, proximity and covariation. *Royal Society open science* **5**, 181100.
- [31] Nunn, C. L., Borgerhoff Mulder, M. & Langley, S. 2006 Comparative methods for studying cultural trait evolution: A simulation study. *Cross-Cultural Research* **40**, 177-209.
- [32] Currie, T. E., Greenhill, S. J. & Mace, R. 2010 Is horizontal transmission really a problem for phylogenetic comparative methods? A simulation study using continuous cultural traits. *Philosophical Transactions of the Royal Society B-Biological Sciences*. **365**, 3903–3912.
- [33] Nunn, C. L., Arnold, C., Matthews, L. J. & Borgerhoff Mulder, M. 2010 Simulating Trait Evolution for Cross-Cultural Comparison. *Philosophical Transactions of the Royal Society B. Biological Sciences*. **365**, 3807- 3819.
- [34] Gray, R. D., Greenhill, S. J. & Ross, R. M. 2007 The pleasures and perils of Darwinizing culture (with phylogenies). *Biological Theory* **2**, 360-375.
- [35] Rogers, A. R. & Cashdan, E. 1997 The phylogenetic approach to comparing human populations. *Evolution and Human Behavior* **18**, 353-358.
- [36] Borgerhoff Mulder, M. 2001 Using phylogenetically based comparative methods in anthropology: More questions than answers. *Evolutionary Anthropology* **10**, 99-111.
- [37] Freckleton, R. P. & Jetz, W. 2009 Space versus phylogeny: disentangling phylogenetic and spatial signals in comparative data. *Proceedings of Royal Society (London) B*. **276**, 21-30.
- [38] Hrnčír, V., Duda, P., Šaffa, G., Květina, P. & Zrzavý, J. 2020 Identifying post-marital residence patterns in prehistory: A phylogenetic comparative analysis of dwelling size. *PLOS ONE* **15**, e0229363. (DOI:10.1371/journal.pone.0229363).
- [39] Sellen, D. W. & Hruschka, D. J. 2004 Extracted-food resource-defense polygyny in native western North American Societies. *Current Anthropology* **45**, 707-715.
- [40] Cashdan, E. & Rogers, A. R. 1997 Book review of "Human nature: A critical reader". *Evolution and Human Behavior* **18**, 279-283.
- [41] Teixidor-Toneu, I., Kool, A., Greenhill, S. J., Kjesrud, K., Sandstedt, J. J., Manzanilla, V. & Jordan, F. M. this issue Historical, archaeological, and linguistic evidence test the phylogenetic inference of Viking-Age plant use. *Phil. Trans. R. Soc Lond. B*.
- [42] Currie, T. E., Meade, A., Guillon, M. & Mace, R. 2013 Cultural phylogeography of the Bantu Languages of sub-Saharan Africa. *Proceedings of the Royal Society B: Biological Sciences* **280**, 20130695.
- [43] Tëmkin, I. & Eldredge, N. 2007 Phylogenetics and material cultural evolution. *Current Anthropology* **48**, 146-153.
- [44] Burton, M. L., Moore, C. C., Whiting, J. W. M. & Romney, A. K. 1996 Regions based on social structure. *Current Anthropology* **37**, 87-123.

- [45] Currie, T. E. & Mace, R. 2014 Evolution of cultural traits occurs at similar relative rates in different world regions. *Proceedings of the Royal Society B* **281(1795)**, 20141622.
- [46] Rogers, D. S. & Ehrlich, P. R. 2008 Natural selection and cultural rates of change. *Proceedings of the National Academy of Sciences* **105**, 3416-3420. (DOI:doi/10.1073/pnas.0711802105).
- [47] Jordan, P. 2014 *Technology as Human Social Tradition: Cultural Transmission among Hunter Gatherers*. Berkeley, University of California Press.
- [48] Royston, P., Altman, D. G. & Sauerbrei, W. 2006 Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in medicine* **25**, 127-141.
- [49] Moravec, J. C., Atkinson, Q., Bowern, C., Greenhill, S. J., Jordan, F. M., Ross, R. M., Gray, R., Marsland, S. & Cox, M. P. 2018 Post-marital residence patterns show lineage-specific evolution. *Evolution and Human Behavior* **39**, 594-601. (DOI:10.1016/j.evolhumbehav.2018.06.002).
- [50] Surowiec, A., Snyder, K. T. & Creanza, N. 2019 A worldwide view of matriliney: using cross-cultural analyses to shed light on human kinship systems *Phil. Trans. R. Soc. B*, 20180077. (DOI:<https://doi.org/10.1098/rstb.2018.0077>).
- [51] Borgerhoff Mulder, M. 1995 Bridewealth and its correlates: Quantifying changes over time. *Current Anthropology* **36**, 573-603.
- [52] Caldwell, J. C., Reddy, P. H. & Caldwell, P. 1983 The causes of marriage change in South India. *Population studies* **37**, 343-361.
- [53] Macfarlan, S. J., Quinlan, R. J. & Post, E. 2019 Emergent matriliney in a matrifocal, patrilineal population: a male coalitionary perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20180073. (DOI:doi:10.1098/rstb.2018.0073).
- [54] Shenk, M. K., Begley, R. O., Nolin, D. A. & Swiatek, A. 2019 When does matriliney fail? The frequencies and causes of transitions to and from matriliney estimated from a de novo coding of a cross-cultural sample. *Philosophical Transactions of the Royal Society B* **374**, 20190006.
- [55] Collard, M., Shennan, S. J. & Tehrani, J. J. 2006 Branching versus blending in macroscale cultural evolution: A comparative study. In *Mapping our Ancestors: Phylogenetic Approaches in Anthropology and Prehistory* (eds. C. P. Lipo, M. J. O'Brien, M. Collard & S. J. Shennan). New York, Aldine Transaction.
- [56] Atkinson, Q. D., Coomber, T., Passmore, S., Greenhill, S. J. & Kushnick, G. 2016 Cultural and Environmental Predictors of Pre-European Deforestation on Pacific Islands. *PLOS ONE*. (DOI:<http://dx.doi.org/10.1371/journal.pone.0156340>).
- [57] Pagel, M., Atkinson, Q. D. & Meade, A. 2007 Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* **449**, 717-720.
- [58] Griffiths, P. E. & Neumann-Held, E. M. 1999 The many faces of the gene. *Bioscience* **49**, 656-662.
- [59] Turchin, P., Currie, T. E., Whitehouse, H., François, P., Feeney, K., Mullins, D., Hoyer, D., Collins, C., Grohmann, S., Savage, P., et al. 2018 Quantitative historical analysis uncovers a single dimension of complexity that structures global variation in human social organization. *PNAS* **115**, E144-E151. (DOI:10.1073/pnas.1708800115 %J Proceedings of the National Academy of Sciences).
- [60] Whitehouse, H., François, P., Savage, P. E., Currie, T. E., Feeney, K. C., Cioni, E., Purcell, R., Ross, R. M., Larson, J. & Baines, J. 2019 Complex societies precede moralizing gods throughout world history. *Nature* **568**, 226-229.

- [61] Miranda, L. & Freeman, J. 2020 The two types of society: Computationally revealing recurrent social formations and their evolutionary trajectories. *PLOS ONE* **15**, e0232609. (DOI:10.1371/journal.pone.0232609).
- [62] Shin, J., Price, M. H., Wolpert, D. H., Shima, H., Tracey, B. & Kohler, T. A. 2020 Scale and information-processing thresholds in Holocene social evolution. *Nature Communications* **11**, 2394. (DOI:10.1038/s41467-020-16035-9).
- [63] Kappeler, P. M., Clutton-Brock, T., Shultz, S. & Lukas, D. 2019 Social complexity: patterns, processes, and evolution. (Springer).
- [64] Sanderson, M. J., Purvis, A. & Henze, C. 1998 Phylogenetic supertrees: assembling the trees of life. *Trends in Ecology & Evolution* **13**, 105-109.
- [65] Duda, P. & Zrzavý, J. 2016 Human population history revealed by a supertree approach. *Scientific reports* **6**, 29890.
- [66] Fortunato, L. & Jordan, F. 2010 Your place or mine? A phylogenetic comparative analysis of marital residence in Indo-European and Austronesian societies. *Phil. Trans. R. Soc. B.* **365**, 3913–3922. (DOI:doi:10.1098/rstb.2010.00173913).
- [67] Ross, C., Strimling, P., Ericksen, K., Lindenfors, P. & Borgerhoff Mulder, M. 2016 The Origins and Maintenance of Female Genital Modification across Africa. *Human Nature* **27**, 173-200. (DOI:DOI 10.1007/s12110-015-9244-5).
- [68] Butler, M. A. & King, A. A. 2004 Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. *American Naturalist* **164**, 683–695.
- [69] Minocher, R., Duda, P. & Jaeggi, A. V. J. E. 2018 Explaining marriage patterns in a globally representative sample through socio-ecology and population history: A Bayesian phylogenetic analysis using a new supertree of human cultures. *Evolution and Human Behavior*.
- [70] Boettiger, C., Coop, G. & Ralph, P. 2012 Is your phylogeny informative? Measuring the power of comparative methods. *Evolution: International Journal of Organic Evolution* **66**, 2240-2251.
- [71] Pagel, M., Meade, A. & Barker, D. 2004 Bayesian estimation of ancestral character states on phylogenies. *Systematic biology* **53**, 673-684.
- [72] Holden, C. J., Meade, A. & Pagel, M. 2005 Comparison of maximum parsimony and Bayesian Bantu language trees. In *The Evolution of Cultural Diversity* (eds. R. Mace, C. J. Holden & S. Shennan), pp. 53-66. London, UCL Press.
- [73] Greenhill, S. J., Currie, T. E. & Gray, R. D. 2009 Does horizontal transmission invalidate cultural phylogenies? . *Proceeding of the Royal Society B-Biol. Sci.* **276**, 2299-2306.
- [74] Walker, R. S., Hill, K. R., Flinn, M. V. & Ellsworth, R. M. 2011 Evolutionary History of Hunter-Gatherer Marriage Practices. *PlosONE* **6**, e19066. (DOI:doi:10.1371/journal.pone.0019066).
- [75] Hage, P. & Marck, J. 2011 Proto-Bantu descent groups. Kinship, Language, and Prehistory. In *Kinship, Language, and Prehistory: Per Hage and the Renaissance in Kinship Studies* (eds. D. Jones & M. Bojka), pp. 75–78. Salt Lake City, University of Utah Press.
- [76] Puttick, M. N., Ingram, T., Clarke, M. & Thomas, G. H. 2020 MOTMOT: Models of trait macroevolution on trees (an update). *Methods in Ecology and Evolution* **11**, 464-471.
- [77] Cooper, N., Thomas, G. H., Venditti, C., Meade, A. & Freckleton, R. P. 2016 A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biological Journal of the Linnean Society* **118**, 64-77.
- [78] Jorgensen, J. G. 1980 *Western Indians. Comparative Environments, Languages, and Cultures of 172 Western American Indian Tribes*. San Francisco, Freeman.

- [79] Bouckaert, R. R., Bower, C. & Atkinson, Q. D. 2018 The origin and expansion of Pama–Nyungan languages across Australia. *Nature Ecology & Evolution* **2**, 741-749.
- [80] Ruhlen, M. 1987 *A Guide to the World's Languages*. 2nd edition ed. CA, Stanford University Press.
- [81] Mathew, S. & Perreault, C. 2015 Behavioural variation in 172 small-scale societies indicates that social learning is the main mode of human adaptation. *Proceedings of the Royal Society B: Biological Sciences* **282**, 20150061.
- [82] Revell, L. J. 2012 phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**, 217-223.
- [83] Paradis, E. & Schliep, K. 2019 ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526-528.
- [84] Pagel, M. & Meade, A. 2006 Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *American Naturalist* **167**, 808-825.
- [85] Beaulieu, J. M. & O'Meara, B. 2019 OUwie: Analysis of Evolutionary Rates in an OU Framework. R package version 1.57. (DOI:<https://CRAN.R-project.org/package=OUwie>).
- [86] Hadfield, J. D. 2010 MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software* **33**, 1-22.
- [87] Botero, C. A., Gardner, B., Kirby, K. R., Bulbulia, J., Gavin, M. C. & Gray, R. D. 2014 The ecology of religious beliefs. *Proceedings of the National Academy of Sciences* **111**, 16784-16789.
- [88] Evans, C. L., Greenhill, S. J., Watts, J., Botero, C. A., Gray, R. D. & Kirby, K. R. this issue The Uses and Abuses of Tree Thinking in Cultural Evolution.

S1) WNAI phylogeny

The figure shows the phylogenetic relationships among the 172 Western North American societies. The colors reflect whether societies and their ancestors lived in forests (red) or not (black).

S2) Pama-Nyungan phylogeny

The figure shows the phylogenetic relationships among the 306 Western Pama-Nyungan societies. The colors reflect whether societies and their ancestors are food-producers (red) or hunter-gatherers (black).

S3) Example of property of the trait horizontal transmission within clade

Horizontal transmission of states within clades can homogenize societies within a clade, enhancing differences among societies in different socioecologies and clades.

error rates discrete trait

<u>direction of change</u>	<u>socioecological influences</u>		<u>lineage differences</u>	
Type 1: false positive	Type 1: false positive		Type 1: false positive	
WNAI PamaNy	WNAI	PamaNy	WNAI	PamaNy
+74% +58%	+33%	+33%	+7%	+9%
Type 2: false negative	Type 2: false negative		Type 2: false negative	
WNAI PamaNy	WNAI	PamaNy	WNAI	PamaNy
+36% -21%	+4%	+/-0%	-13%	-18%

error rates continuous trait

<u>direction of change</u>	<u>socioecological influences</u>		<u>lineage differences</u>	
Type 1: false positive	Type 1: false positive		Type 1: false positive	
WNAI PamaNy	WNAI	PamaNy	WNAI	PamaNy
+70% +38%	+/-0%	+44%	+57%	+36%
Type 2: false negative	Type 2: false negative		Type 2: false negative	
WNAI PamaNy	WNAI	PamaNy	WNAI	PamaNy
-41% -26%	+99%	+80%	+/-0%	+/- 0%

S4) Example of property of the trait

horizontal transmission from geographic neighbor

Geographic neighbours are often not the same as phylogenetic neighbours, which can introduce divergent states into clades, biasing the inference of internal nodes.

error rates discrete trait

<u>direction of change</u>	<u>ecological influences</u>		<u>lineage differences</u>	
Type 1: false positive	Type 1: false positive		Type 1: false positive	
WNAI PamaNy	WNAI	PamaNy	WNAI	PamaNy
+45% +35%	-10%	-18%	-10%	-8%
Type 2: false negative	Type 2: false negative		Type 2: false negative	
WNAI PamaNy	WNAI	PamaNy	WNAI	PamaNy
+19% -21%	+3%	+6%	-25%	-3%

error rates continuous trait

<u>direction of change</u>	<u>ecological influences</u>		<u>lineage differences</u>	
Type 1: false positive	Type 1: false positive		Type 1: false positive	
WNAI PamaNy	WNAI	PamaNy	WNAI	PamaNy
+87% +98%	+7%	+36%	+82%	+29%
Type 2: false negative	Type 2: false negative		Type 2: false negative	
WNAI PamaNy	WNAI	PamaNy	WNAI	PamaNy
-42% -51%	+89%	+89%	+/-0%	+/- 0%

S5) Example of property of the trait

fewer variants of the trait

Grouping variants together into single categories (e.g. all shades of grey classified as black) reduces power, potentially suggesting stronger differences among lineages

error rates discrete trait

<u>direction of change</u>	<u>ecological influences</u>		<u>lineage differences</u>	
Type 1: false positive	Type 1: false positive		Type 1: false positive	
WNAI PamaNy	WNAI	PamaNy	WNAI	PamaNy
-3% -2%	-1%	+/-0%	-2%	+10%
Type 2: false negative	Type 2: false negative		Type 2: false negative	
WNAI PamaNy	WNAI	PamaNy	WNAI	PamaNy
+28% +18%	-2%	+3%	+6%	-1%