Ceramic diversity, random copying, and tests for selectivity in ceramic production

James Steele*, Claudia Glatz, Anne Kandler

AHRC Centre for the Evolution of Cultural Diversity, Institute of Archaeology, University College London, 31–34 Gordon Square, London WC1H 0PY, UK

ARTICLE INFO

ABSTRACT

This paper examines the evidence for selective reproduction of ceramic types based on their evolving frequency distribution, and on correlations of the abundance of types with underlying functional characteristics. How can we tell if a variant has been preferentially selected or discriminated against, and become more or less frequent than would be expected under pure drift (i.e. random copying, with the compounding over time of proportionate sampling error)? We examined a database of Hittite ceramic bowl types from two successive Phases of occupation of the Upper City of Boğazköy–Hattusa, capital of the Hittite empire and the largest Bronze Age settlement in Turkey. We applied two tests used in genetics to assess departures from neutrality, to assess selectivity in rates of reproduction of ceramic bowl types (the Ewens-Watterson and Slatkin's Exact tests). We also examined the effects of ceramic fabric and vessel dimensions on changes in vessel abundance between the two Phases, using regression analysis. We found that while the frequency distribution of rim sherds did not in itself enable us to reject the null hypothesis of random copying, closer examination of the characteristics of these types enabled us to recognize latent dimensions of functional variability (including ware type and bowl diameter) that had demonstrably been the subject of selective decision-making by the potters. The present case study suggests that we should be wary of applying the neutral model from genetics uncritically in archaeology, because it is much harder to prove that the cultural traits whose frequencies are being modelled are genuinely functionally equivalent (as that model requires).

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

This paper examines the evidence for selective reproduction of ceramic types based on their evolving frequency distribution, and on correlations of the abundance of types with underlying functional characteristics. It is intended as a contribution to evolutionary archaeology.

It is a truism that cultural traditions and successful innovations are socially transmitted both between and within generations (respectively, by vertical or oblique and by horizontal transmission routes; Cavalli-Sforza and Feldman, 1981), with learners applying heuristics or rules of thumb in choosing when to engage in independent trial-and-error learning, and in selecting which models to copy when this is the preferred strategy (transmission biases; Boyd and Richerson, 1985). Boyd and Richerson (1985) distinguish several different decision mechanisms affecting the adoption of new cultural traits, each of which is determined by an underlying expectation that people seek to copy the most advantageous variants.

For quantitative hypothesis testing, we also need a random copying model as a null hypothesis, if we want to recognize the outcomes of non-random (selective) copying. The neutral model of cultural diversity in finite populations in the absence of selection is one in which new traits appear by a process akin to random mutation, and existing traits are randomly chosen to be copied so that their frequencies change only because of chance sampling effects. Numerous recent archaeological studies have examined the stylistic choices made by potters and their clients from such a theoretical perspective. Following Neiman's (1995) work on cultural transmission of selectively neutral stylistic elements in Woodland ceramics, which introduced neutral theory from genetics and explored evidence for changes in the scale of interaction and in the number of potters whose vessels were being circulated in a particular study region, more recent archaeological applications of cultural transmission theory have concentrated on the influence of social norms on the rates of transmission of particular cultural variants (Shennan and Wilkinson, 2001; Kohler et al., 2004; Eerkens and Lipo, 2005; Schauer, 2008).

We will not present a detailed analysis of the neutral model here, since this model has already been expounded for archaeologists at greater length by Neiman (1995) and by Shennan and Wilkinson (2001). Briefly, Eq. (1) summarizes the expected cultural

* Corresponding author.
  E-mail address: j.steele@ucl.ac.uk (J. Steele).

homogeneity F within a population at a time t under this model, as a function of the homogeneity in the previous time step (F_{t-1}), the number of individuals producing successive generations of cultural variants (N_e, with a generation meaning an independent episode of cultural production at which variation may be introduced), and the mutation rate μ (Neiman 1995, p. 14, Eq. (2)):

$$F_t = \left[\frac{1}{N_e} + \left(1 - \frac{1}{N_e}\right)F_{t-1}\right](1-\mu)^2 \qquad (1)$$

The rate of mutation μ (a constant between 0 and 1) is the probability that an individual copying episode introduces a new variant rather than faithfully copying an existing one. The likelihood of an individual making a faithful copy of an existing model is $(1-\mu)$. The probability that any two randomly sampled individual copying episodes at time t have faithfully copied the same exemplar from the previous copying round t_{-1} is $1/N_e$, while $1 - 1/N_e$ represents the probability that they have not. Equilibrium homogeneity $\widehat{F}_{eq}$ is reached when the rate of mutation corresponds to that of random variant loss ($F_t = F_{t-1}$), since loss and mutation are opposing evolutionary forces, the former reducing diversity, the latter increasing it (Shennan and Wilkinson, 2001, p. 583):

$$\widehat{F}_{eq} = \frac{(1-\mu)^2}{\left[N_e - (N_e - 1)(1-\mu)^2\right]} \qquad (2)$$

For transmission systems in which the rate of mutation μ is very low, this can be approximated as:

$$\widehat{F}_{eq} \approx \frac{1}{2N_e\mu + 1} \qquad (3)$$

which indicates that the equilibrium homogeneity of neutral variants within a population is inversely proportional to $2N_e\mu$.

We emphasize that to apply this model to cultural production we must assume one of two simplifications of reality. On the one hand, we might assume that each human producer always produces identical artifacts (in which case a generation is a human biological generation, a copying episode is an episode of vertical or oblique transmission of craft skills between a skilled and an unskilled individual, and N_e is the number of active teachers). On the other hand, we might assume that a single human producer can produce many successive cultural generations of variants, and that at each successive copying episode an individual is no more likely to copy a variant in his or her own repertoire than an extant variant found elsewhere in the population (in which case a generation is an independent bout of craft production, a copying episode may involve inter-individual transmission or may simply involve the same individual reproducing artifacts in a new bout of production, and N_e is the number of active producers). In this paper, we make the second of these assumptions; however, we note that future developments of this model could usefully address the contrasting empirical rates of mutation and of sampling of alternative variants in within-individual versus between-individual copying episodes.

Fig. 1 shows typical output from a simulation of evolving ceramic assemblage diversity under drift. In this case, we start with a uniform distribution of 10 variants, each making up 10% of the assemblage, and then propagate them over a hundred copying cycles using the random copying rule, with new variants introduced with a likelihood $\mu = 0.01$. As this example shows, it is obvious that under drift most variants dwindle towards extinction while a few others increase in abundance, and that the diversity profile evolves towards one in which the empirical homogeneity statistic $F = \sum_{i=1}^{k} p_i^2$ fluctuates around the value expected at mutation–drift equilibrium ($\widehat{F}_{eq} \approx 1/2N_e\mu + 1$).

Archaeological use of this model has focused on two issues: the inference of changes over time in the values of N_e or μ (respectively the size of the population reproducing the variants, and the mutation rate); and inference of non-neutrality (i.e., the existence of other processes than drift influencing variant frequencies) at a given location and time step. Where there is assumed to have been no departure from neutrality, variation in the size of the population reproducing the variants and/or in the mutation rate has been inferred from the frequency distribution of variants. Neiman (1995) found a trend across time for an increase in the value of an index of diversity, θ, in inventories of Woodland ceramic stylistic attributes, which he interpreted as reflecting an increase in the scale of the social network and in the number of pottery producers. This index of diversity can be calculated empirically as the coefficient $\theta = (1/F) - 1$, and can also be estimated as a maximum likelihood for an observed sample size and number of variants using Ewens' sampling formula (Ewens, 1972). Instead of calculating θ, Bentley et al. (2004) have used the magnitude of the power law slope coefficient α to make similar inferences about varying innovation rates in other cultural contexts where the population size is known. They draw on another technique for estimating how well an empirical assemblage matches the predictions of the neutral model by plotting the frequency distribution of variants, with the expectation under neutrality of a linear fit on a log–log plot of trait frequency against the likelihood of a trait appearing at that frequency (which means that the underlying distribution can be fitted as a first approximation by a power-law curve, cf. Hahn and Bentley, 2003; Bentley et al., 2004; Herzog et al., 2004). This derives from Kimura and Crow's (1964) demonstration that the equilibrium homogeneity $\widehat{F}_{eq} = 1/2N_e\mu + 1$ corresponds to a frequency distribution of neutral variants with expected number of variants at frequency v, $P(v) = \theta v^{-1}(1-v)^{\theta-1}$, which approximately corresponds to a power law distribution $P(v) \sim v^{-\alpha}$ for small mutation rates μ (with α varying as a function of θ; Hahn and Bentley, 2003; Bentley et al., 2004).

How can we tell if a variant has been preferentially selected or discriminated against, and become more or less frequent than would be expected under pure drift? One approach, popular in both population genetics and in ecology during the early phases of development of neutral theory, has been to test for departures from the theoretical frequency distribution of variants that is predicted by the neutral model. This is the approach explicitly used by Lipo (2001), and which is implicit in the studies by Shennan and Wlikinson (2001) and by Kohler et al. (2004). In genetics, two such tests use Ewens' (1972) sampling formula. This formula predicts the properties of the underlying distribution for a neutral process, where an empirical sample of n cases has been observed with k variants. The Ewens–Watterson test of homozygosity (Watterson, 1977, 1978) evaluates the empirical homogeneity statistic $F = \sum_{i=1}^{k} p_i^2$ against the value expected under neutrality, and Watterson (1978) provides a table of values for F for different values of n (up to 500 observations) and k (up to 10 variants) and for different significance thresholds. This is similar to the approach of Shennan and Wilkinson (2001) except that the latter did not estimate statistical significance. The Slatkin Exact Test evaluates the empirical frequency distribution against all possible configurations for a given n and k under the Ewens sampling distribution, and is a more general test which makes fewer assumptions about which aspects of distribution shape might indicate selection (Slatkin, 1994, 1996). Slatkin (1997) has helpfully archived online his source code for a Monte Carlo simulation program that conducts both tests (Ewens–Watterson and Slatkin Exact). The outputs are tail probabilities, which indicate the position of an empirical set of data in a probability distribution of possible sets derived from the Ewens
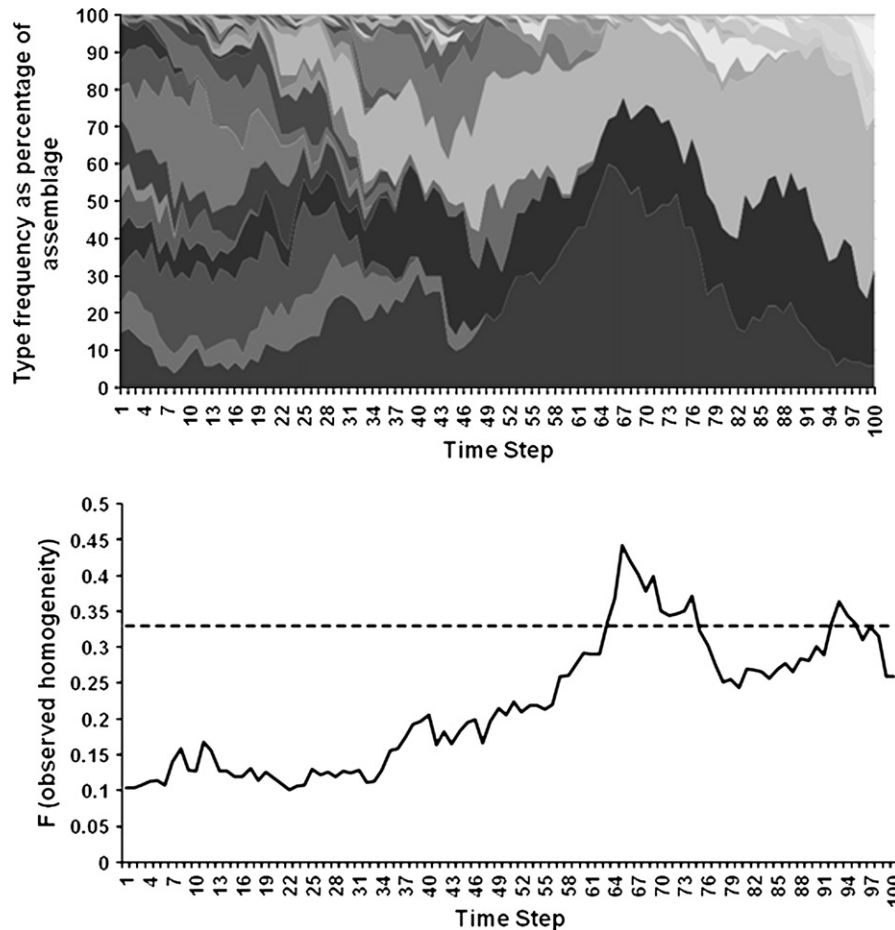
Fig. 1. A sample simulation of the effects of drift on assemblage diversity. The initial condition is ten types each making up 10% of the assemblage, which then evolves under random copying with a constant population size of $N_e = 100$ independent copying events in each cycle, and a mutation rate $\mu = 0.01$. The upper graph shows the changing proportions of the assemblage made up by each type (types are vertically stacked in order of appearance over the time course of the simulation). The theoretically expected equilibrium homogeneity level F is 0.33, shown as the dotted line in the lower graph, which also shows the empirical homogeneity level evolving towards that mutation–drift equilibrium over the time course of the simulation.

sampling formula. P_H is the tail probability for the Ewens–Watterson homozygosity test and P_E is that for the Slatkin Exact test. To evaluate departures from neutrality using a two-tailed test, we might select a significance level α (e.g. 0.05) and then assert that the null 'drift' model is rejected by one or the other test if P_H or P_E is either less than $\alpha/2$ or greater than $(1 - \alpha/2)$. If the tail probability for either test is lower than $\alpha/2$ it indicates significantly greater evenness, and if it is greater than $(1 - \alpha/2)$ it indicates significantly greater concentration than that expected from a neutral process. If we do not want to set a critical α value, then we can simply report the *p*-values without imposing an arbitrary significance level (cf. Gigerenzer, 2004).

In both genetics and ecology, however, strong tests of departure from neutrality (and thus of the presence of selection) have been shown to involve rather more than simply testing the goodness of fit of empirical data to a theoretical frequency distribution (e.g. Kreitman, 2000; Ford, 2002; Wootton, 2005; McGill et al., 2006; Harpole and Tilman, 2006). In reality, the shape of the empirical frequency distribution – even when found to depart significantly from the theoretical shape – need not indicate departures from neutrality if the system is not at equilibrium. If, for instance, the innovation rate μ has recently changed, or if N_e has recently changed, or if the population is incompletely mixing, then there may be an excess of variants at certain frequencies without any departure from neutrality in the underlying general process (e.g.

Slatkin, 1982; Watterson, 1986). Frequency-based measures are also poor at detecting episodes of positive or negative selective sweeps on a fairly neutral background, because of the transient nature of such effects (Zhai et al., 2009). In addition to statistical tests like the Ewens–Watterson and the Slatkin Exact, stronger tests are therefore needed which examine possible effects of trait characteristics on variants' abundance. The simplest form of such a test for cultural traits would be to analyze statistically the effects of some parameter on a variant's abundance – for example, the effects of design complexity on abundance for decorative stylistic traits on pots, where the initial hypothesis might be that an effect is expected because of the extra time and cognitive load required to produce the more complex variants. For variants which cannot be assumed to be functionally equivalent, but where one wishes nonetheless to evaluate the possibility that their relative abundance reflects random drift, the test would involve some clearly functionally-relevant parameter [for example, a neutral model of registration rates of different pedigree dog breeds (Herzog et al., 2004) could be tested by looking for effects of breed characteristics on registration rate, since breeds range enormously in size, temperament, and suitability to different tasks and environments]. If the empirical data can be broken down into successive time steps, then a further hypothesis can be advanced about the effects of such non-equivalent traits on the changing representation of variants over time.

Finally, a separate prediction for the evolution of cultural traits in a random copying condition relates to variation in metrical attributes of artefacts being copied. Based on the Weber–Fechner law in psychophysics, which implies that humans cannot by their own unaided senses reliably discriminate differences in some scalar perceptual variable of less than $\pm n\%$ (estimated empirically to be of the order of 5%, Eerkens, 2000), we can predict that dimensions of some artefact type being copied in a transmission chain with no selective constraint will vary in the manner of a geometric random walk. This means that an ensemble of artefacts each the product of its own transmission chain will evolve from an initially uniform condition to an asymptotic lognormal distribution of metrical variation in any selectively-neutral dimension, with the median value decreasing and the variance increasing as a function of time. In terms of its drift component this is an identical argument to that of the neutral model for the evolution of frequencies of discrete traits, although here we assume no effects of mutation. The application of such a neutral model to metrical attributes of archaeological artefacts is due to Eerkens and Lipo (Eerkens, 2000; Eerkens and Lipo, 2005); its implications have recently been explored in relation to changing sizes of archaeological stone tools by Hamilton and Buchanan (2009). An appropriate set of tests for such variables would examine the goodness-of-fit of the empirical size distribution to a lognormal model, and also examine the direction of change over time in the sample medians and variances. This gives us an additional test of selectivity in cultural reproduction of ceramic attributes by examining change in metrical attributes of an ensemble of types over time, and we shall explore its relevance in the case study analysed below.

2. Materials and methods

In this paper, we shall examine a database of Hittite ceramic bowl types from two successive Phases of occupation of the Upper City of Boğazköy-Hattusa, capital of the Hittite empire and the largest Bronze Age settlement in Turkey, which lies around 150 km east of the modern Turkish capital of Ankara on the central Anatolian plateau. The ca. 180 ha site occupies a rocky terrain and slopes around 300 m from south to north. Occupation in the lower part (Unterstadt) of the Hittite capital stretches back to the third millennium B.C. During the Late Bronze Age, a large temple, storage facilities and living quarters dominated the Lower City. The Upper City (Oberstadt) is a ca. 100 ha southerly addition to the older part of the town and the palatial area on Büyükkale. The area was first enclosed by 3.4 km long fortification wall in the second part of the LBA and includes 30 monumental structures or temples, numerous smaller buildings, a pottery production area in the second Phase, water reservoirs and a monumental glacis for defensive as well as representational purposes. The majority of the excavated ceramic material from the temple quarter in the central Upper City at Boğazköy-Hattusa derives from the final two LBA occupation Phases (Oberstadt 3 and Oberstadt 2, referred to here as O.St.3 and O.St.2). These are distinguished architecturally by a Phase of abandonment at the end of O.St.3 and a partial stratigraphic overlap of later O.St.2 structures. Houses and monumental buildings belonging to O.St.3 appear to have been abandoned and left to decay, some were leveled and new functions assigned to the freed-up spaces. House 4, for instance, appears to have been abandoned and the ground leveled to make way for two phases of pottery kilns (Müller-Karpe, 1988, p. 5). The Upper City material in O.St.2 is the only unequivocal case of pottery production that was at least physically attached to monumental structures. At the end of the Late Bronze Age the temples and other structures were abandoned and some were destroyed by fire (Seeher, 2001). The entire Upper City complex was initially dated by archaeologists to the last

decades of the city's LBA existence (late 13th century B.C.) (Müller-Karpe, 1988, p. 161; Parzinger and Sanz, 1992. pp. 72–73). The original, textually driven chronology of Neve (1992, 1999) would allow as little as 50 years for the two main occupation phases. More recent archaeological work and a re-assessment of the chronological span of tablet and glyptic archives, however, suggest a much longer occupation of the Upper City from at least the 15th century B.C. onwards (Mielke et al., 2006).

The Late Bronze Age ceramic tradition is firmly rooted in the preceding Middle Bronze Age, with the majority of vessel shapes being produced for around 600 years (ca. 1800–1200 B.C.). Few shapes appear or fall entirely out of use over the course of its history, and consequently trends within that period have been recognizable only from frequency seriation of stratified assemblages (Schoop, 2003, 2006). At the time of writing, this frequency seriation has been defined quantitatively only for the second part of the LBA based on analysis of the two latest LBA occupation Phases – O.St.3 and O.St.2, identified by independent archaeological criteria – of the Upper City at Boğazköy-Hattusa, (Müller-Karpe, 1988; Parzinger and Sanz, 1992; also Schoop, 2003, 2006). In this paper, we concentrate on the published ceramic material from the Upper City. With regard to the Upper City pottery assemblages, Müller-Karpe (1988, pp. 161–162) and Parzinger and Sanz (1992, p. 47) have pointed out a number of changes over time. These include changes in the predominant formal characteristics between O.St.3 and O.St.2 such as increasingly thickened rims. They also noted a replacement of popular vessel forms by shapes previously less frequently attested. For instance, bowls with inverted rims appear to have been replaced by bowls with everting rims, while the role of plates with stepped rim profiles seems to have been taken over by shallow bowls or platters with simple rounded rims. Müller–Karpe and Parzinger and Sanz also observed a general shift from a more even distribution of plain coarse and finer wares in the first of the two occupation phases, with additionally the occurrence of red/brown and white slipped vessels, to a much more uniform utilization of mostly coarse plain wares across the entire Upper City ceramic spectrum. O.St.2 is conventionally seen as the final Phase of decline in the NCA ceramic tradition at the very end of the LBA, in terms both of formal diversity and of the use of finer wares and attention to surface treatment.

In the absence of decorative motifs, the null hypothesis of selectively neutral transmission will be tested using frequencies of discrete artefact types. To avoid clearly functionally contingent variation in vessel frequencies, we will be using formal variation (rim shapes) within a single functional category (bowls). It may be that there is still functionally-significant diversity within this category, and the various tests used here – based both on overall frequency distributions of variants, and on the physical characteristics of each variant in relation to its abundance – will enable us to evaluate whether or not this is the case. Like most pre-existing typologies in the archaeological literature, the Parzinger and Sanz NCA pottery typology is based on a hierarchical taxonomy that is extensionally derived, although types are defined in some descriptive detail. We are aware of the limitations of using such a typology to reconstruct transmission processes. Evolutionary archaeologists have pointed out that hierarchical taxonomic classification systems obscure the variation between individual instances which is key to understanding transmission dynamics. These scholars have advocated a paradigmatic classification, which is better suited to take account of individual artefact variation (Neff, 1993; O'Brien and Lyman, 2002; see Sterling, 2001 for a study of Egyptian Meidum bowls based on a paradigmatic classification). However, in this case the very detailed variant categorization defined by Parzinger and Sanz is likely to capture even rare variants

in rim shape, while their database also includes records of variation in bowl diameter and in fabric and surface coating within each of these rim shape-based groups and variants (Figs. 2 and 3).

The Parzinger and Sanz (1992) dataset includes a total of 6927 contextually secure rim fragments. Parzinger and Sanz (1992) identified 17 (A–P, Q for bases) functional vessel categories (jars, cooking pots, bowls, etc.). These 17 vessel categories are further subdivided into a total of 60 formal groups and 244 sub-groups. The catalogue lists finds context, vessel type and formal classification, ware type, number of rim pieces as well as estimated
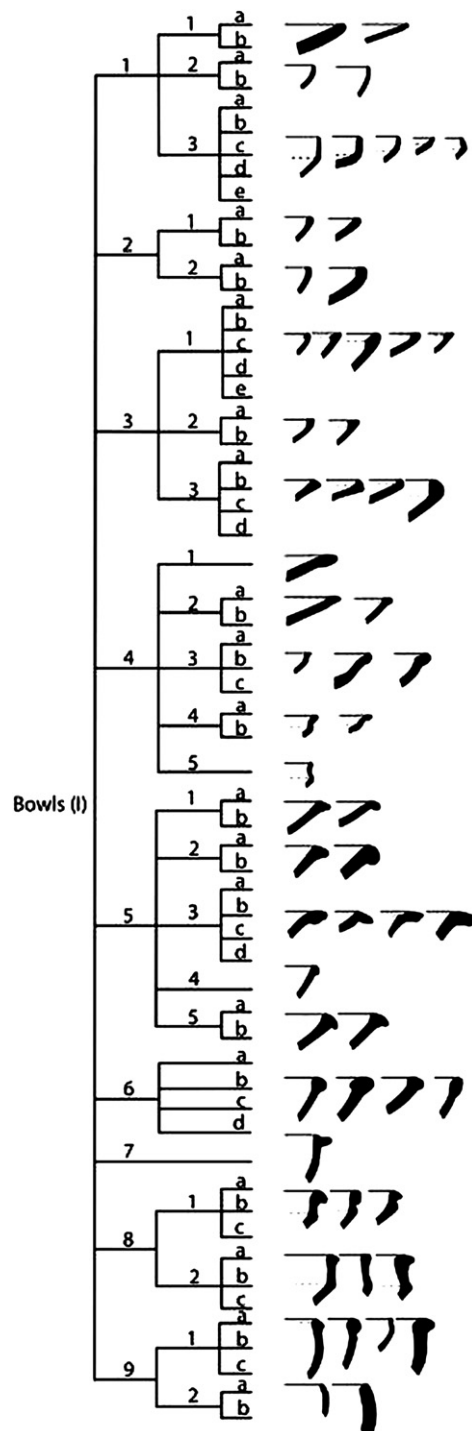
rim diameters. Bowl fragments are numerically the most prominent vessel category in the Upper City assemblage. Parzinger and Sanz (1992; see Fig. 4) identified nine primary formal bowl groups: I-1 (bowls with simple rounded rims), I-2 (bowls with simple thickened rim), I-3 (bowls with inverted rim), I-4 (bowls with sinuous profiles and everted rims), I-5, I-6, I-7 (bowls with everted rims), I-8 (carinated bowls with everted rims) and I-9 (bowls with inverting walls). Each of these nine groups is further subdivided into one to five formal sub-groups, and up to five variants within each of the sub-groups, yielding a total of 61 different forms at the variant level. Of these we have excluded the four variants of subtype I5.3 from our analyses, since the rim shapes in this group (three sherds from O.St. 3 and 11 sherds from O.St. 2) belong to jars rather than bowls. Relatively large sample sizes (O.St.3: 1392, and O.St.2: 2061) as well as the wide range of formal types and variants make the bowl assemblages a promising starting point for testing the neutral model. We shall analyse changing frequencies of bowl types at two taxonomic levels (main bowl groups, and variants), as the main bowl groups might be expected to define any functionally significant variation in rim form while the variants might be expected to be more subject to drift.

We want to model changing vessel frequencies in this study, on the basis of counts of rim sherds. The Parzinger and Sanz (1992) data tabulation consists primarily of rim fragments, but does not report the percentage of the rim circumference represented by each sherd. This prevents the estimation of vessel equivalents (Estimated Vessel Equivalent, EVE; Orton et al. 1993, pp. 166–175) as a measure of actual vessel frequency. As a crude first check for any systematic bias of vessel size on sherd count, we can however test for correlations between rim sherd frequencies and bowl diameters in our sample. Fig. 4 plots average bowl diameters against rim sherd frequencies for the 57 bowl variants in both Phases. It is obvious that there is no significant overall correlation between bowl diameter and rim sherd frequency in this assemblage. Other things being equal, larger vessels tend to break into more pieces than smaller ones. However, larger vessels in this assemblage are also more likely to be made of coarse plain wares (A), while finer wares (E) and surface slips (C, D) are generally more often associated with smaller rim diameters. The greater fragility of these small vessels may have contributed to the observed lack of a bowl size/sherd count relationship. Nonetheless, even if there was no systematic bias towards greater numbers of rim sherds per vessel for particular main bowl groups or variants, rim sherd counts may still inflate the overall sample size for the exact tests for neutrality (which require counts of independent episodes of individual vessel production), since a single vessel may be represented by more than one rim sherd. Such sample size inflation could give rise to Type 1 errors (where the null hypothesis is mistakenly rejected), although it would not be expected to give rise to Type 2 errors. In cases where the exact tests give a marginally significant result for departures from neutrality, we will therefore flag this as an issue.

The Late Bronze Age wheel-made pottery from the Upper City is classified according to four overall ware categories (A, C, D, E; see Table 1). Variation in color, temper, firing atmosphere, and hardness are expressed in sub-categories (e.g. A4b). For the purpose of this study, we used only the four overall ware distinctions. Ware A consists of red/brown medium to coarse mineral tempered clay (0.5–3 mm inclusions) with little or no surface treatments beyond smoothing and occasional washes. Ware C combines a finer fabric with grain sizes rarely exceeding 0.5 mm and fabric color ranging from beige to red with red/brown polished slips. Ware D consists of fine as well as coarser fabric types (0.2–3 mm inclusions) with beige to white slips of varying thickness. Ware E designates a fine
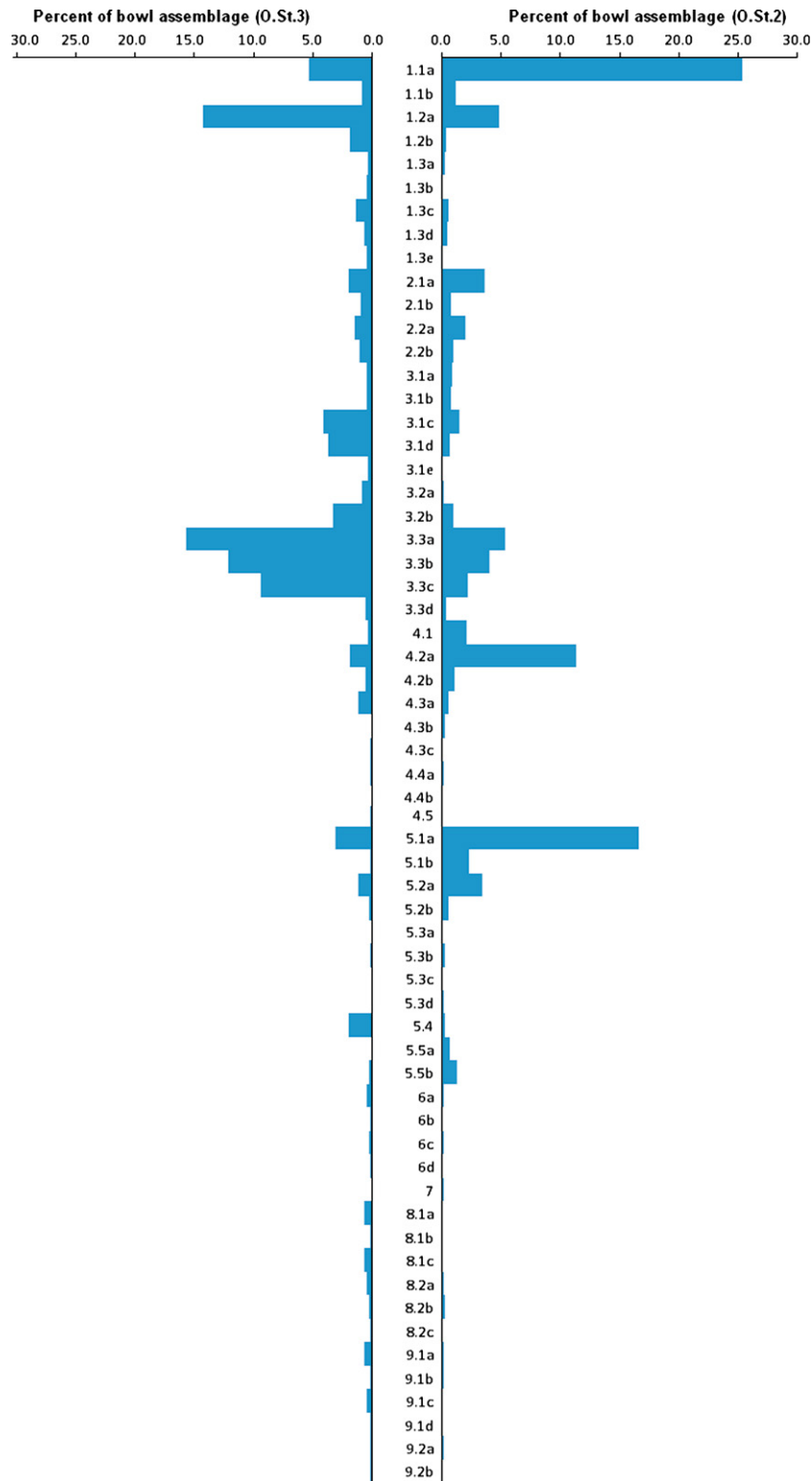

Fig. 2. Upper City ceramic bowl repertoire (modified after Parzinger and Sanz, 1992).

Fig. 3. Percentages of each bowl type in the Upper City rim sherd assemblages from O.St.3 and O.St.2.

beige to red fabric (0.1–1 mm inclusions) with smoothed or plain surfaces.

Data were analysed in Microsoft Excel 2003 (Microsoft Corporation, 2003) and in SPSS 14.0 for Windows (SPSS Inc., 2005), and the Ewens-Watterson and Slatkin Exact tests were run using Slatkin's Monte Carlo program (Slatkin, 1997) compiled in Microsoft Visual C++ Version 6.0 (Microsoft Corporation, 1998). Power law exponents were calculated from log–log regressions of rank against abundance, following Adamic (1999). The expression for a rank-abundance curve of power law form is $n \sim r^{-b}$, where r is the rank and n is the abundance, and b is a constant. This is equivalent to a cumulative distribution function of the form $r = n^{-1/b}$ (given that r
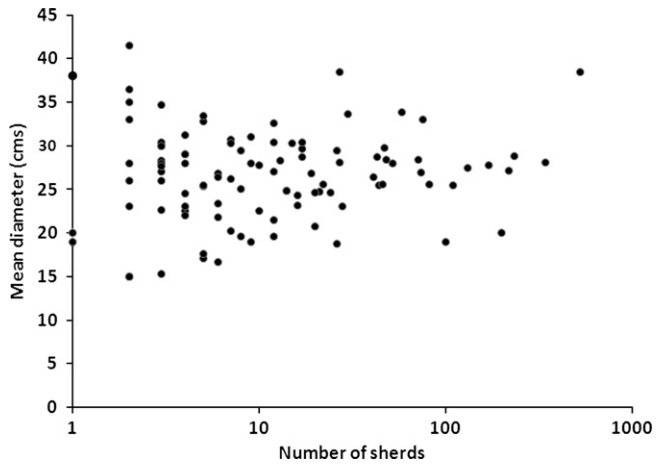
Fig. 4. Scattergram of average bowl diameter against rim sherd frequency, for the 57 bowl variants analysed. Average estimated bowl diameters may vary slightly between phases for variants, so values are plotted for each Phase separately.

is also the number of occurrences of size $\geq n$), from which we obtain the exponent $\alpha = (1 + 1/b)$ for the power law probability density function $P(v) = v^{-\alpha}$ (Adamic, 1999).

3. Results

The ranked frequency distributions of main bowl groups and of variants are plotted in Fig. 5. We first consider the test statistics relating to these frequency distributions from each Phase (see Table 2). The absolute values of the diversity indices θ are of course greater for the fine-grained typology (variants) than for the main bowl types, because F will inevitably be smaller when each variant makes up a relatively small proportion of the overall assemblage. For the main bowl groups, the marginally statistically significant value for P_E in the O.St. 3 assemblage reflects the fact that for this Phase there are no main bowl groups with very low rim sherd

frequencies (the Slatkin Exact test is more sensitive to what is going among the rarer variants). This suggests balancing selection among bowl types that had different functions or contexts of use, although we need to exercise caution as the rim sherd corpus, as a record of fragment frequency, will tend to include multiple records of a single vessel and may therefore have inflated the sample size sufficiently to induce a Type 1 error. The Ewens–Watterson Test, on the other hand, does not flag this main bowl assemblage as departing significantly from neutrality. Neither test flags the bowl assemblage from the later Phase (O.St.2) as departing significantly from neutrality in its main bowl group counts, nor in its counts of variants.

These results, which suggest the absence of selective biases on bowl assemblage composition in the later Phase, are however contradicted by analysis of effects of type characteristics on frequencies. Some initial insight may be gained by considering the main bowl groups whose abundance changes most markedly between the two Phases. Table 1 gives the frequencies for each of the nine main bowl groups in each Phase. It is immediately apparent that in terms of absolute abundance the bowls with inverted rims have more than halved their representation, the bowls with simple rounded or thickened rims have nearly doubled their representation, and the bowls with everted rims have nearly quadrupled their representation. This confirms Müller–Karpe's and Parzinger and Sanz's comments on an apparent change-over of dominant bowl types from bowls with inverted rims in O.St.3, to bowls with everted rims and simple straight rims in O.St.2. In addition, the least frequent main bowl groups (I-6 to I-9) have more than halved their abundance.

At the lower taxonomic level of bowl variants, a linear regression analysis with stepwise variable selection (Table 3) showed that the abundance rank of bowl variants in the later phase was predicted not only by their rank in the earlier phase (as would be expected under drift), but also by the proportion of each variant that had been made using the plain coarse ware in that earlier phase. There was no additional effect of vessel diameter or of percentage of variants made in the other ware types in the earlier phase (these variables were excluded by the stepwise regression

Table 1
Frequencies of rim sherds of the main bowl groups in each of the four wares, by phase.

|	Ware A (plain coarse)	Ware C (red slip)	Ware D (white slip)	Ware E (plain finer)	Total
O.St.3					
Type I1 (bowls with simple rounded rims)	80	111	28	141	360
Type I2 (bowls with simple thickened rims)	22	12	8	36	78
Type I3 (bowls with inverted rims)	171	276	53	214	714
Type I4 (bowls with everted rims)	22	19	5	19	65
Type I5 (bowls with everted rims)	40	16	5	37	98
Type I6 (bowls with everted rims)	7	2	1	6	16
Type I7 (bowls with everted rims)	–	–	–	–	–
Type I8 (carinated bowls with everted rims)	5	11	2	17	35
Type I9 (bowls with inverting walls)	6	7	2	11	26
Total	353	454	104	481	1392
	25.4%	32.6%	7.5%	34.5%	
O.St.2					
Type I1 (bowls with simple rounded rims)	590	32	3	61	686
Type I2 (bowls with simple thickened rims)	94	5	–	52	151
Type I3 (bowls with inverted rims)	240	35	4	68	347
Type I4 (bowls with everted rims)	300	8	1	15	324
Type I5 (bowls with everted rims)	501	6	1	11	519
Type I6 (bowls with everted rims)	5	–	–	3	8
Type I7 (bowls with everted rims)	2	1	–	1	4
Type I8 (carinated bowls with everted rims)	4	–	–	7	11
Type I9 (bowls with inverting walls)	11	–	–	–	11
Total	1747	87	9	218	2061
	84.8%	4.2%	0.4%	10.6%	
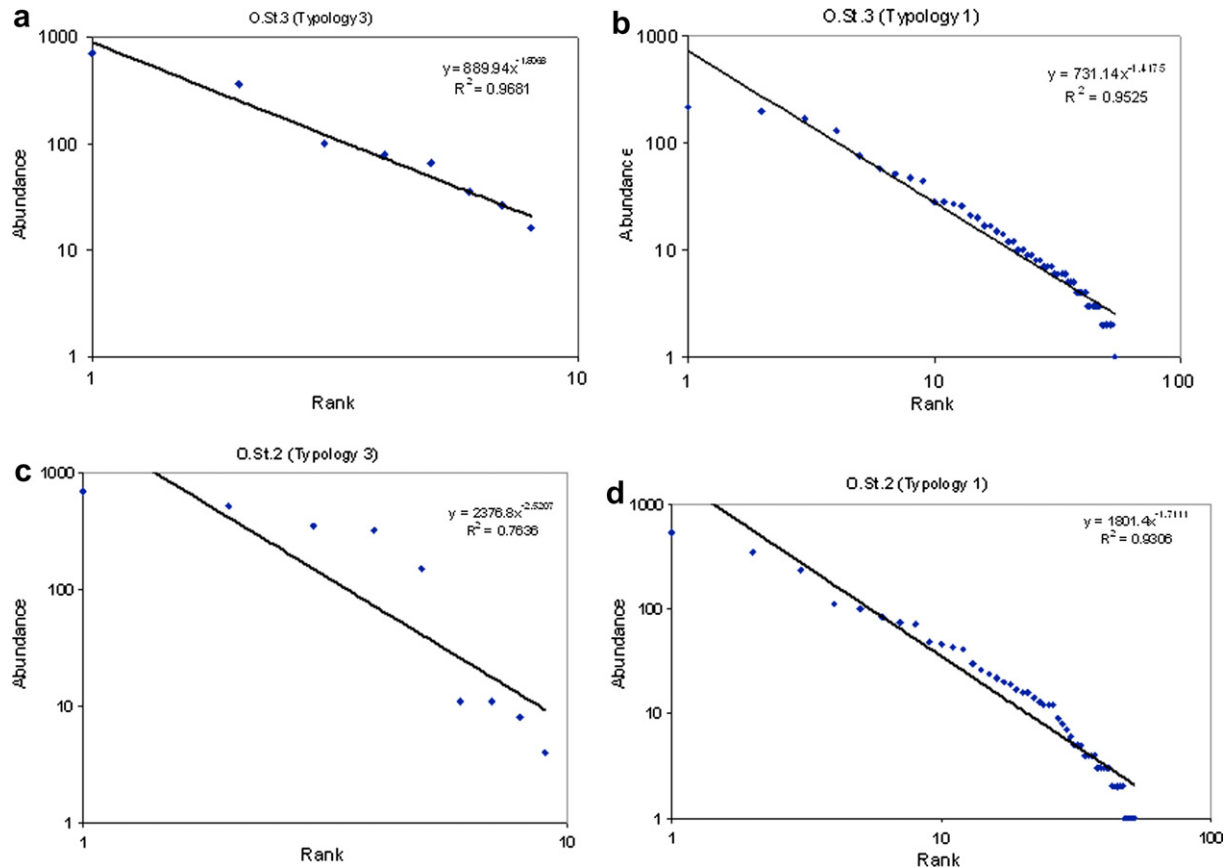
Fig. 5. (a) Rank-abundance plot (both axes logged) for the O.St.3 ceramic assemblage ($n = 1392$ sherds, eight main bowl groups). (b) Rank-abundance plot (both axes logged) for the O.St.3 ceramic assemblage ($n = 1392$ sherds, 54 variants). (c) Rank-abundance plot (both axes logged) for the O.St.2 ceramic assemblage ($n = 2061$ sherds, nine main bowl groups). (d) Rank-abundance plot (both axes logged) for the O.St.2 ceramic assemblage ($n = 2061$ sherds, 52 variants).

analysis). Bowl shapes that were produced more often in plain coarse ware in the first phase tended to become more popular in the second phase. This is inconsistent with a neutral model and indicates that by the later phase, popularity of bowl types was associated with functionally-significant characteristics that had become subject to selective decision-making by the potters or their clients.

Ware frequencies are also summarized for each phase in Table 1. The most common wares in O.St.3 are the finer plain ware, which makes up 34.5% of rim sherds, and the red slip-coated ware, which makes up 32.7% of bowl rim sherds. In the later phase, these percentages decline to 10.6% and 4.3% respectively. Correlation analysis of the bowl variants in O.St.3 found no significant correlation between variants' rank in or percentage of the assemblage, and either mean diameter or the proportion of the variant made in

Table 2
Statistical measures of the frequency distributions of the main bowl groups and variants.

Phase	Level of typology	n	k	F	θ_F	θ_E	P_H	P_E
O.St. 3	Main	1392	8	0.3414	1.93	1.030	0.256	0.0218
O.St. 3	Variant	1392	54	0.0803	11.45	11.04	0.569	0.050
O.St. 2	Main	2061	9	0.2327	3.30	1.122	0.042	0.101
O.St. 2	Variant	2061	52	0.1179	7.48	9.578	0.841	0.382

n, Number of rim sherds; k, number of variants. $F = \sum_{i=1}^{k} p_i^2$. $\theta_F = (1/F) - 1$. $\theta_E = \theta$ estimated from Ewens' (1972) sampling formula $E_k = \sum_{i=0}^{n-1} \theta_E/(\theta_E + i)$. P_H, tail probability for Ewens–Watterson homozygosity test; P_E, tail probability for Slatkin's Exact test.

coarse (or any other) ware. The data on wares in Table 1 suggest strongly that there was selective reproduction of large, coarse plain ware vessels in the later phase (O.St.2). We have already seen that by the later phase (O.St.2; Table 3), bowl variants made in coarse plain ware had become subject to positive cultural selection. Bowls with larger diameters in the earlier phase (O.St.3) tended also to be made more often in the coarse plain fabric (mean diameter versus percent made in coarse plain ware, $n = 53$ O.St.3 bowl variants, Spearman's rho $= 0.32$, $p = 0.02$). Fig. 6 shows a trend not apparent from the variant frequencies themselves, which is for the bowl variants which increase in relative frequency (higher rank) in O.St.2 also to be of increased mean diameter in that later Phase while retaining the rim form that is the basis of their typological classification (change in rank versus change in mean diameter, $n = 48$ O.St.2 bowl variants, Spearman's rho $= 0.46$, $p = 0.001$). The most likely reason for this is that bowls produced in the coarse fabric in O.St.3 tended also to be among the larger realizations of that variant, so that when we find in O.St.2 a larger proportion of that variant made in coarse fabric the mean diameter is also greater. The direction of change in the overall distribution of bowl sizes also appears to be inconsistent with a neutral copying model in metric dimensions, since the mean diameters increased over time (mean bowl diameter in O.St.3 = 26.5 cm, in O.St.2 = 29.7 cm) and the distribution is better fitted by a normal than a lognormal model; moreover the later sample has proportionally fewer cases in the left-hand tail of the size distribution. All of these observations indicate a trend in the opposite direction to that predicted under neutral copying with compounded and proportionate error.

Table 3

Linear regression analysis (with stepwise variable selection) of the effects of bowl variant characteristics in the earlier phase (O.St.3) on abundance in the later phase (O.St.2).

Dependent Variable	Predictor	B Coefficient (Unstandardised)	t	Sig.	Pearson's r
Rank of variant in O.St.2	Constant	16.6 ± 3.8	4.4	<0.001	0.71
	Rank of variant in O.St.3	0.64 ± 0.1	5.8	<0.001	
	Percentage of variant made in plain coarse ware (A) in O.St.3	-25.1 ± 7.6	-3.3	0.002	
Variant as % of total O.St.2 assemblage	Constant	-0.01 ± 0.01	-0.6	0.56	0.46
	Variant as % of total O.St.3 assemblage	0.49 ± 0.16	3.0	0.004	
	Percentage of variant made in plain coarse ware (A) in O.St.3	0.05 ± 0.03	2.2	0.034	

Relative abundance of bowl variants in O.St.2 is predicted by their relative abundance in the previous phase (O.St.3), and also by the percentage of the variant that was made in coarse plain ware in that earlier phase (for rank the latter correlation is negative, because rank gets smaller as a variant is proportionally more abundant; thus the most abundant variant has a Rank of 1, etc.). There is no separate effect of average vessel diameter or of percentage of variants made in the other ware types in the earlier phase (these variables were excluded by the stepwise regression analysis).

Thus, the bowl variants in O.St.3 vary in frequency independently of fabric or diameter. The frequency of sherds from the main bowl groups in that earlier Phase suggests some balancing selection, although the frequency distribution of variants does not depart from that expected in a 'random drift' process. However, superimposed on a drift-like signature, in the later Phase (O.St.2) we also found selective reproduction of bowl forms with simple rounded or thickened rims and with everted rims, with large diameters (and these tend to have become larger), and made from coarse fabric. This would not have been detected from the Exact tests' results for within-phase diversity based on counts of rim sherds.

Fig. 6. (a) Proportional frequencies of bowls of all types, by diameter and phase. Mean bowl diameter in O.St.3 = 26.5 cm, in O.St.2 = 29.7 cm. (b) Correlation between changing abundance ranking and change in average diameter of bowl variants represented in both phases (O.St.3 and O.St.2). Pearson's $r = 0.40$, $p = 0.006$.

4. Discussion

In some recent attempts to formulate a theory of cultural dynamics, models were adopted fairly directly from biology, often without much modification to take account of the differences which characterize cultural transmission as an inheritance system (for instance, the non-random nature of human trial-and-error strategies when exploring a technological design space, for the importance of which see e.g. O'Brien and Mesoudi, 2008a, b; and the high prevalence of oblique and horizontal transmission, biased by inductive heuristics that increase the likelihood of learners identifying selectively advantageous traits). These differences do not invalidate the application of broader Darwinian principles to cultural transmission systems, but they do require that new models be developed to explore micro- and macro-evolutionary cultural dynamics. The present case study suggests that we should be wary of applying the neutral model from genetics uncritically in archaeology, because it is much harder to prove that the cultural traits whose frequencies are being modelled are genuinely functionally equivalent (as that model requires). This reminds us that inference of cultural microevolutionary processes from archaeological remains is very difficult, for all the reasons usually associated with inverse problems (e.g. Boyd and Richerson, 2008). In assessing the fit between a model and a set of data, forward approaches to modelling use the known dynamics of the empirical system to predict outcomes for a given parameter constellation. In inverse problems, the outcomes are known to some degree, but the dynamics of the empirical system and the parameter constellation are unknown and must be estimated by 'reverse engineering'. Typically in such situations, difficulties arise where parameter values cannot be reliably estimated from observable data, and where it can be shown that alternative models and alternative parameter constellations would yield the same observed outcomes.

In the case study analysed here, we did not have direct access to the micro-scale decisions of potters in replicating individual types, and the frequency distribution of rim sherds did not in itself enable us to reject the null hypothesis of random copying. However, closer examination of the characteristics of these types enabled us to recognize latent dimensions of functional variability (including ware type and bowl diameter) that had demonstrably been the subject of selective decision-making by the potters. Our results are also consistent with conclusions drawn by other analysts on less explicitly quantitative grounds. If we follow Müller-Karpe's (1988) functional interpretation, the changes in vessel type frequencies between the two Phases of occupation of the Upper City of Boğazköy–Hattusa suggest an increased importance in food consumption/preparation for firm over liquid or semi-liquid food, a preference for larger vessels, and an increased preference for coarse and potentially less labor-intensive ware types. The long-term trend of decreasing numbers of elaborate slipped and
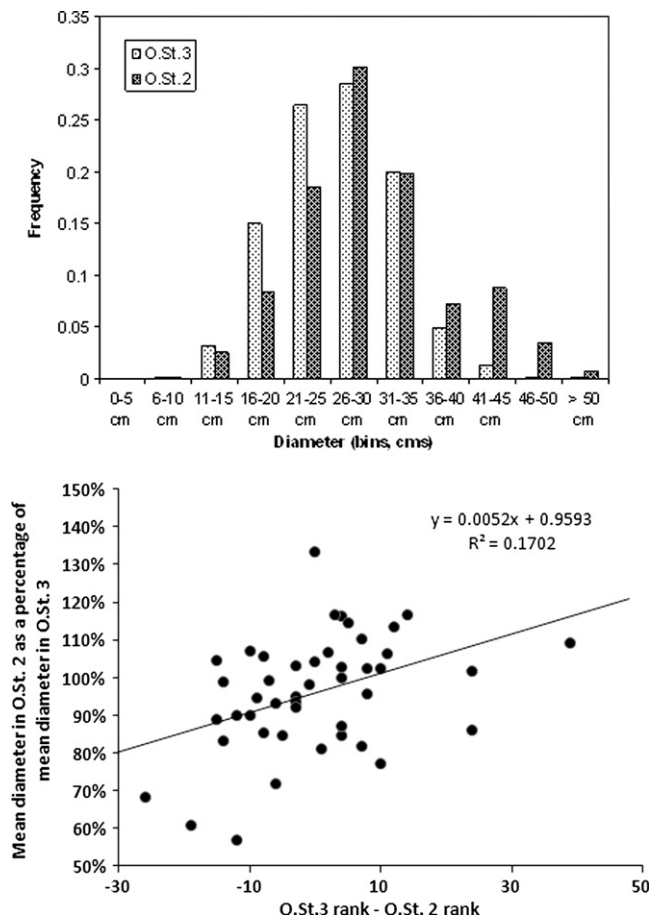
polished vessels may be related to the decreasing role of high-investment pottery in social contests. Late Bronze Age Near Eastern and East Mediterranean value systems, at least at the top of the social hierarchy, appear to have moved on to other materials such as precious metals, stone or ivory for social and political negotiation. In the case of shallow bowls with simple rounded rims (I1.1.) which become a dominant type in the O.St.2 repertoire we may be seeing a functional amalgamation or change-over from previously popular plates with stepped rims (Müller-Karpe, 1988, p. 127), for which frequent traces of secondary burning indicate a function in food preparation (Schoop, 2003, p. 173).

In the present study the traits analysed were formal typological ones, some of which potentially affect bowl function (for example, the presence or absence of inverted rims). Analyzing traits in a neutral model when their functional equivalence is not proven is a familiar strategy in ecology. A neutral model has been applied there as a null model of species abundance (for example, to explain the abundance ranking of tropical forest tree species at a given location), and here the presumption of functional equivalence among species is highly debatable (e.g. Hubbell, 2005). Nonetheless the neutral model has been hard to falsify, and its application has led to the recognition that species abundance distributions may reflect multiplicative random processes as well as adaptive variation in niche characteristics. Furthermore, strong tests of departure from neutrality (and thus of the presence of selection) have been shown to involve rather more than simply testing the goodness-of-fit of empirical data to a theoretical frequency distribution. In this paper we have used such a strong test, and by showing a correlation between vessel characteristics and their abundance ranking in the later Phase (O.St.2) we have identified clear evidence of departure from neutrality (the neutral model assumes functional vessel equivalence, and would therefore predict the absence of any such correlation). There is considerable literature on the association of vessel performance in, for instance, cooking tasks or the retention of liquids with fabric coarseness, types of inclusions, wall-thickness and surface treatment (Rice, 1987). At LBA Boğazköy-Hattusa, selection was primarily on the ware types and perhaps secondarily on the bowl dimensions, and there seems also to have been selection against the inverted rim forms. The exact tests for neutrality were conducted on rim sherds, the morphology of which bears only an indirect and uncertain relationship to some of these functionally-significant properties. Not all the lower-order rim variants need necessarily have been functionally significant in their own right. Some of them may have increased in frequency through 'hitchhiking', if they were already associated with a predominantly coarse ware realization.

There are exciting prospects for an integrated science of cultural dynamics which has robust models of the population dynamics of copying processes, as well as good inferential controls on the units of transmission in such processes. Analysis of this database also reminds us of the difficulty of inferring, from the fragmented sample of archaeologically-recovered artefact forms and attributes, the ideational units of cultural transmission which our copying models require us to identify. Rim sherd counts give only first approximations of the underlying variation in vessel type frequency, while vessel types themselves are archaeological constructs, which at best should be seen as working hypotheses about the boundaries between artefact categories that had meaning for their creators. Fabric analysis may also demonstrate that wares fall on a continuously-varying spectrum, rather than into a set of discrete underlying classes, while vessel size estimation from rim fragments is an inexact science. These caveats should not prevent us from trying to characterize assemblage diversity using statistical techniques, but they remind us that even as we continue to develop the methods and models of the kind outlined in this paper, we must also pay attention to refining our understanding of what features and categories can be treated as reliable units of prehistoric cultural transmission.

References

Adamic, L.A., 1999. Zipf, power-laws, and pareto – a ranking tutorial. Available from: <http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html> [accessed May 27, 2008].

Bentley, A.R., Hahn, M.W., Shennan, S.J., 2004. Random drift and culture change. Proceedings of the Royal Society 271, 1443–1456.

Boyd, R., Richerson, P.J., 1985. Culture and the Evolutionary Process. Chicago University Press, Chicago.

Boyd, R., Richerson, P.J., 2008. Response to our critics. Biology and Philosophy 23, 301–315.

Cavalli-Sforza, L.L., Feldman, M.W., 1981. Cultural Transmission and Evolution: A Quantitative Approach. Princeton University Press, Guildford.

Eerkens, J.W., 2000. Practice makes within 5% of perfect: the role of visual perception, motor skills, and human memory in artifact variation and standardization. Current Anthropology 41 (4), 663–668.

Eerkens, J.W., Lipo, C.P., 2005. Cultural transmission, copying errors, and the generation of variation in material Culture and the archaeological record. Journal of Anthropological Archaeology 24, 316–334.

Ewens, W., 1972. The sampling theory of selectively neutral alleles. Theoretical Population Biology 3, 87–112.

Ford, M.J., 2002. Applications of selective neutrality tests to molecular ecology. Molecular Ecology 11, 1245–1262.

Gigerenzer, G., 2004. Mindless statistics. Journal of Socio-Economics 33, 587–606.

Hahn, M.W., Bentley, R.A., 2003. Drift as a mechanism for cultural change: an example from baby names. Proceedings of the Royal Society London B 270, S120–S123.

Hamilton, M.J., Buchanan, B., 2009. The accumulation of stochastic copying errors causes drift in culturally transmitted technologies: quantifying Clovis evolutionary dynamics. Journal of Anthropological Archaeology 28, 55–69.

Harpole, W.S., Tilman, D., 2006. Non-neutral patterns of species abundance in grassland communities. Ecology Letters 9, 15–23.

Herzog, H.A., Bentley, R.A., Hahn, M.W., 2004. Random drift and large shifts in popularity of dog breeds. Proceedings of the Royal Society London B (Suppl.) 271, S353–S356.

Kimura, M., Crow, J.F., 1964. The number of alleles that can be maintained in a finite population. Genetics 49, 725–738.

Kohler, T.A., Van Buskirk, S., Ruscavage-Barz, S., 2004. Vessels and villages: evidence for conformist transmission in early village aggregations on the Pajarito Plateau, New Mexico. Journal of Anthropological Archaeology 23, 100–118.

Kreitman, M., 2000. Methods to detect selection in populations with applications to the human. Annual Review of Genomics and Human Genetics 1, 539–559.

Lipo, C.P., 2001. Science, style, and the study of community structure: an example from the Central Mississippi River Valley. BAR International Series 918. Archaeopress, Oxford.

McGill, B.J., Maurer, B.A., Weiser, M.D., 2006. Empirical evaluation of neutral theory. Ecology 87, 1411–1423.

Mesoudi, A., O'Brien, M.J., 2008a. The cultural transmission of Great Basin projectile-point technology I: an experimental simulation. American Antiquity 73, 3–28.

Mesoudi, A., O'Brien, M.J., 2008b. The cultural transmission of Great Basin projectile-point technology II: an agent-based computer simulation. American Antiquity 73, 627–644.

Microsoft Corporation, 1998. Microsoft Visual C++ 6.0. Microsoft Corporation.

Microsoft Corporation, 2003. Microsoft Office Excel 2003. Microsoft Corporation.

Mielke, D.P., Schoop, U.-D., Seeher, J. (Eds.), 2006. Strukturierung und Datierung der hethitischen Archäologie: Voraussetzungen – Probleme – Neue Ansätze. Internationaler Workshop Istanbul, 26–27. November 2004. BYZAS 4. Eğe Yayınları, Istanbul.

Müller-Karpe, A., 1988. Hethitische Töpferei der Oberstadt von Hattuša. Ein Beitrag zur Kenntnis spät-großreichszeitlicher Keramik und Töpferbetriebe unter Zugrundelegung der Grabungsergebnisse von 1978–1982 in Boğazköy. Hitzeroth, Marburg.

Neff, H., 1993. Theory, sampling and analytical techniques in the archaeological study of prehistoric ceramics. American Antiquity 58, 23–44.

Neiman, F.D., 1995. Stylistic variation in evolutionary perspective: inferences from decorative diversity and inter-assemblage distance in Illinois Woodland ceramic assemblages. American Antiquity 60, 7–36.

Neve, P., 1992. Hattuša, Stadt der Götter und Tempel. Neue Ausgrabungen in der Hauptstadt der Hethiter. Philipp von Zabern, Mainz.

Neve, P., 1999. Die Oberstadt von Hattuša, Die Bauwerke I, Das Zentrale Tempelviertel, Boğazköy-Hattuša, vol. XVI. Gebr. Mann, Berlin.

O'Brien, M.J., Lyman, R.L., 2002. The epistemological nature of archaeological units. Anthropological Theory 2, 37–56.

Orton, C., Tyers, P., Vince, A., 1993. Pottery in Archaeology. Cambridge Universtiy Press, Cambridge.

Parzinger, H., Sanz, R., 1992. Die Oberstadt Keramik von Hattuša. Hethitische Keramik aus dem Zentralen Tempelviertel. Gebr. Mann, Berlin.

Rice, P.M., 1987. Pottery Analysis: A Sourcebook. University of Chicago Press, Chicago.

Schauer, P., 2008. Cultural evolution in the age of Athens: drift and selection in Greek figure-painted pottery. Ph.D. Thesis, University College London.

Schoop, U.-D., 2003. Pottery traditions of the later Hittite empire: problems of definition. In: Fischer, Bettina, Genz, Hermann, Jean, Éric, Köroğlu, Kemalettin (Eds.), Identifying Changes: The Transition from the Bronze to the Iron Ages in Anatolia and its Neighbouring Regions. Proceedings of the International Workshop Istanbul, November 8–9, 2002. Türk Eskiçağ Bilimleri Enstitüsü, Istanbul, pp. 176–178.

Schoop, U.-D., 2006. Dating the Hittites with statistics. Ten pottery assemblages from Boğazköy-Hattuša. In Strukturierung und Datierung der hethitischen Archäologie: Voraussetzungen – Probleme – Neue Ansätze. Internationaler Workshop Istanbul, 26–27 November 2004, edited by Dirk P. Mielke, Ulf.-Dietrich Schoop and Jürgen Seeher, pp. 215–239. BYZAS 4. Istanbul: Eğe Yayınları.

Seeher, J., 2001. Die Zerstörung von Hattuša. In Akten des IV. In: Wilhelm, Gernot (Ed.), Internationalen Kongresses für Hethitologie, Würzburg, 4–8 Oktober 1999. Studien zu den Boğazköy-Texten 45. Harrassowitz, Wiesbaden, pp. 623–634.

Shennan, S.J., Wilkinson, J.R., 2001. Ceramic style and neutral evolution: a case study from Neolithic Europe. American Antiquity 66 (4), 577–593.

Slatkin, M., 1982. Testing neutrality in subdivided populations. Genetics 100, 533–645.

Slatkin, M., 1994. An exact test for neutrality based on the Ewens sampling distribution. Genetic Research 64, 71–74.

Slatkin, M., 1996. A correction to the exact test based on the Ewens sampling distribution. Genetic Research 68, 259–260.

Slatkin, M., 1997. Ewens exact program [use MONTE CARLO] Available from: http://ib.berkeley.edu/labs/slatkin/monty/Ewens_exact.program [accessed 21 July 2008].

SPSS Inc., 2005. SPSS for Windows, Rel. 14.0.1. SPSS Inc., Chicago.

Sterling, S.L., 2001. Social complexity in Ancient Egypt: functional differentiation as reflected in the distribution of standardized ceramics. In: Hunt, Terrence L., Lipo, Carl, P., Sterling, Sarah L. (Eds.), Posing Questions for a Scientific Archaeology. Bergin and Garvey, Westport, pp. 145–173.

Watterson, G.A., 1977. Heterosis or neutrality? Genetics 85, 789–814.

Watterson, G.A., 1978. The homozygosity test of neutrality. Genetics 88, 405–417.

Watterson, G.A., 1986. The homozygosity test after a change in population size. Genetics 112, 899–907.

Wootton, J.T., 2005. Field parameterization and experimental test of the neutral theory of biodiversity. Nature 433, 309–312.

Zhai, W., Nielsen, R., Slatkin, M., 2009. An investigation of the statistical power of neutrality tests based on comparative and population genetic data. Molecular Biology and Evolution 26, 273–283.