RESEARCH ARTICLE

Plainware and Polychrome: Quantifying Perceptual Differences in Ceramic Classification Between Diverse Groups to Further a Strong Objectivity

Lewis Borck*,†, Jan C. Athenstädt‡, Lee Ann Cheromiah§, Leslie D. Aragon‖, Ulrik Brandes¶ and Corinne L. Hofman**

A common problem when classifying archaeological objects is a potential cultural bias of the person deciding on the classification system. These are existing concerns within archaeology and anthropology and have previously been discussed as an emic/etic divide, "folk" classifications, or objective versus subjective approaches. But who gets to decide what is objective is often a subjective endeavour. To examine if and how cultural perceptions bias classification systems, we use methods from the field of cultural domain analysis to quantify differences in perception of ceramic sherds between different groups of people, specifically archaeologists and Indigenous and non-Indigenous potters. For this study, we asked participants to arrange a set of 30 archaeological sherds on a canvas, then interviewed them following each sorting exercise. A geosocial analysis of the arrangements in this pilot study suggests that there are substantial differences in the criteria by which the sherds are sorted between the groups. In particular, the arrangements by the Indigenous potters showed a greater diversity in the selection of underlying attributes. Understanding our different perceptions towards the material we use to construct history is the first step towards approaching a strong objectivity and thus a less fraught and more culturally inclusive discipline.

Keywords: pottery classification; cultural domain analysis; quantification of perception; culturally informed cognition; spatial sorting; strong objectivity

In order to systematize and record objects, to compare sites and to use techniques such as seriation, archaeologists have established classification systems for finds. These classification systems are based on criteria that appear meaningful from the point of view of the archaeologist. While these approaches have advanced the field and have arguably led to a more objective way to study objects, the underlying choices that researchers make as to which factors are important remain subjective since they are based on the culturally informed perception of the archaeologist(s).

In this study, we propose a statistical approach to establish similarities between objects: instead of relying on the perception of a single individual or a small group of archaeologists, we quantify distances and similarities between a set of objects as perceived by a number of different individuals. This allows us to average the perceived distances, which in turn can allow us to detect the features that play an important role in people’s perceptions of ceramics. Moreover, using these distances we can quantify differences in perception between groups, for example between archaeologists and Indigenous potters.

Like the statistical approaches we are using, the research approach we present in this work is of an inductive and exploratory nature rather than one oriented around a hypothetico-deductive model. The methods from cultural domain analysis that we use allow the construction of similarities between objects purely based on the perception of the participants without being constrained by assumptions about the underlying principles. This open approach allows us to capture inter-individual and inter-group differences in order to examine their relationships to the group more holistically. Thus, instead of breaking out individual components, this method allows us to investigate the data relationally.

1. Background

Research with a focus on material culture always sparks debates about how reliable or representative interpretations of past human behaviours and social
processes can be. In any field focused on understanding humanity through their things, and archaeology in particular, this debate is necessary as researchers grapple with understanding contemporary and historical social, religious, and political behaviours through the fracturing prism of the material world. In aggregate, researchers using comparative tools from computer science, the social sciences, and methods developed by archaeologists and anthropologists have advanced our understanding by critically assessing and improving how we study and interpret the material representations of past human (and environmental) behaviour. But this tension between the material record and what it represents, or at least what researchers think it represents, lives on. Take, for instance, ceramic studies within archaeology.

The earliest debates within archaeology, and some of the most vicious, revolved around the analysis of ceramics. Researchers questioned whether typologies reflected real types, were simply constructed units that help create order for comparative analyses, or whether the typologies were even reproducible between different analysts (e.g., Brew 1946; Neff 2006; Rouse 1960; Smith 1979). The Ford-Spaulding debates were perhaps the most visible of this period. In short, Spaulding (1953) argued that quantitative methods, particularly those recently developed by Brainerd (1951) and Robinson (1951), should be used to determine ceramic similarities. Ford’s (1954) position was that ceramic seriation, and thus change, was founded on cultural change and Spaulding’s methodology decidedly missed how cultures evolved.

Typologies were settled on as convenient ways to organize data, even as some continued to recognize the problematic disconnect between biological relationship and cultural relationship. For example, Brew (1946) argued early on that ceramic cladistics that replicated taxonomic conventions developed to examine historic evolutionary relationships in biological organisms was problematic. This critique was expanded by Morris Opler, a cultural anthropologist, as he confronted work by prominent archaeologists such as Kroeber, Ford, and White, and criticized them for essentially studying culture and cultural change while ignoring that human beings, more importantly individuals, were creating that change. As he said while criticizing Ford, “man has developed his enormous and intricate brain, his powers to remember and record the past, his abilities to probe the minute and the remote, his capacity for invention, communication, and planning, in order to remain a supermoron fit only to fetch and carry for Mother [cultural] Evolution” (Opler 1963: 902).

The exact process by which taxonomies (or typologies) were constructed for archaeological ceramics varied extensively by region. The Caribbean, one of the early testing grounds for ceramic taxonomy for chronological purposes (Hofman, Hoogland and Van Gijn 2008: 2), used different systems to construct chronological and cultural typologies (e.g., Barbotin 1974; Bullen 1964; Bullen and Bullen 1968; Gauthier 1974; Hoffman 1967; Mattioni and Bullen 1970; Petitjean Roget 1963, 1968, 1970; Pinchon 1952; Rainey 1940; Sears and Sullivan 1978; Winter 1978). Eventually, many of the ceramic classifications in the Caribbean became founded upon Rouse’s contributions (e.g., Rouse 1960, 1972, 1993). Rouse (1960) argued that typological analyses are bifurcated between natural and artificial categories (see also Brew 1946; Neff 1996). In his framework, types constructed by archaeologists are created by selecting modes (groupings of artefact attributes) that they determine to be relevant. The modes that are selected, however, are inherent to the archaeological material, and thus to the culture that created that material. Thus, typology, according to Rouse, starts at the bottom through an analysis of attributes and elements that are inherent and then proceeds to build a typology by moving up in a series of hierarchical steps away from those natural elements to a more encompassing type that is artificially created. A similar exploration of the nature, meaning, and even utility, of ceramic taxonomies was also underway in places like the American Southwest (e.g., McGimsey 1980; Plog 1980).

Following work by researchers like Balfet (1965), archaeologists moved beyond arguments over the organization of material remains into referential categories and to deeper social and economic questions. However, these investigative approaches still required that the data the researcher analysed was meaningful. The data must capture real trends or the analysis is biased at best. This is why Mills and Crown noted (1995: 4) that the construction of typologies and archaeological attempts at understanding ceramic production (their particular social and economic question) were intricately linked.

Typologies are heavily linked to our ability to interpret the past. Recognizing early on that there was a disconnect between how archaeologists interpreted ceramics (i.e. an etic perspective) and how the producers of those ceramics would have interpreted them (i.e. an emic perspective), archaeologists began applying ethnohistory and ethnographic work to attempt to improve interpretations of ceramic function (e.g., Barbotin 1974; Petitjean Roget 1963; Pinchon 1952).

This approach, often called ethnoarchaeology (though see Gould 1968; Kleindienst and Watson 1956), was championed by the ground-breaking Sally and Lewis Binford (e.g., Binford 1968; Binford and Binford 1969) and many of their contemporaries (e.g., Arnold 1980; Hodder 1982; Kramer 1979; Longacre 1974; Parker Pearson 1982; Spurling 1984; Stark 1991). It represents what some see as a “living archaeology” (Shrotiya 2007), although it often overly relies on an idea that contemporary communities are living ancestors in a way that follows similar upstreaming problems found with the direct historical approach (see Stahl 2017 for an extended discussion). As such, many researchers use the direct historic approach as a comparative assessment tool (e.g., Stahl 2017; Wylie 1985) rather than an explanatory method (i.e. middle range theory).

Many of the best known studies in ethnoarchaeology have focused on hunter-gather groups (e.g., Binford and Binford 1969) to better understand faunal and lithic use in the archaeological record. Yet there is a strong contingent of researchers oriented towards sedentary populations. Because of its prominence as a material type in non-mobile groups, these studies often focus on ceramic
production (e.g., Beck 2006; Rice 1999; Sinopoli 1991; Stark 1991), usually to create a middle range theory set (e.g., Binford, Cherry and Torrence 1983; although see Raab and Goodyear 1984) instead of a comparative tool.

While most ethnoarchaeological studies drew, and continue to draw, much needed connections between the archaeological past and the contemporary present, they primarily have focused on understanding both the present and the past as external, objective researchers (e.g., etc). Non-Western perspectives in particular are often holistic and relational (e.g., Cajete 1999), which can seem non-rigorous against the Enlightenment derived drive to segment and isolate in order to understand. Our study follows Cunningham’s (2003) call to see how ethnoarchaeology can complement Indigenous scholarship on the material past and adds to the growing body of literature that demonstrates that holistic and relational analyses are rigorous.

Importantly, the above arguments parallel many similar discussions happening within the discipline of psychology relating to how people construct meaning of the world external to themselves and how differences of interpretations (bias) may emerge from this construction.

For example, Theory of Mind is what we use to analyze, interpret, judge, and infer other people’s behaviors, and internal thoughts based on body language during our daily interactions (Gweon and Saxe 2013). For researchers looking to interpret data, the Theory of Mind highlights that since we cannot directly observe another individual’s thoughts, even in the present, we necessarily must interpret data to understand the thoughts that underly any human action, or in the case of archaeology and materials, the choices (e.g., Borck and Mills 2017), experiences (e.g., Hegmon 2016), and behaviors (e.g., Schiffer 1976) that together with non-purposeful and emergent human behaviour and environmental interactions construct the archaeological record. As interpretations can be problematically open to implicit and explicit biases (see for example Cunningham and MacEachern 2016), this is the area that can be most problematic in the humanities and social sciences. Historical researchers from any position within that continuum would likely agree that understanding how different perspectives impact interpretations is important for historical reconstructions (Borck 2018; Borck and Sanger 2017; Cunningham 2003; Feyerabend 1993; Henry, Angelbeck and Rizvi 2017).

Social psychologists call our constant interpretation of one another’s actions or behaviour, really the choices and thoughts that form that activity, attribution. An attempt to model this activity is attribution theory (Kelley and Michela 1980). Divergent attributions can arise from people’s differing personal histories, cultural backgrounds, political views, gender, and temporal context.

Since how these histories – or axes – of identity intersect informs how we experience our present (e.g., Combahee River Collective 1977; Crenshaw 1989), it is possible to view the individual as a collection of group identities. Following that, it is possible to examine differences in how varying groups of people perceive and experience particular interactions or material objects (as embodied choices and behaviour) by incorporating some of these various identities as well.

In archaeology, in particular, this can be difficult. Researchers are rarely looking at how individuals interact or interpret each other’s pasts or behaviours. Instead, using the material that comprises the archaeological record, they decipher past people’s choices and behaviors, usually in aggregate, through the material record. Archaeological analysis can therefore lead to heavy distortions when attributions further bias human thought, experience, and behavior already obscured by the material record.

This article examines how various groups construct categories, not just from the raw elements of attributes encoded within individual pieces of ceramics, but also from their cultural background and individual histories. We propose that these intersectionally dependent processes of categorical construction, which we measure using geosocial tests produced from cognitive placement tests of ceramic sherds, can be testable and quantifiable.

2. Analysing Differential Perceptions

In the field of cultural anthropology, cultural domain analysis (CDA) describes a set of methods with the goal to understand the semantic structure of cultural domains, i.e., the mental categories people construct regarding a set of words, images, or other items (Borgatti 1994; Borgatti and Halgin 1999). It is a way of analysing how groups of people create relationships between objects or ideas.

These methods have traditionally been applied in situations where perceptions of groups of people were studied or compared. One strength of these methods is that they allow researchers to measure not only information about perceived attributes of an item (monadic data) but also the perceived relationships between items (dyadic data).

We now briefly introduce the techniques for recording perception that we use in this work. During all experiments, it was crucial to tell the participant that there was no right or wrong answer, but that the goal was to find out their opinion on the relationships between the items.

2.1. Rank Order

In a rank order task (Borgatti 1994; Stephenson 1935), participants are asked to establish an ordering of the items based on a specific attribute. Examples for the attribute could be, how “beautiful” a certain object is or how likely it is to ask a person for a favor (if the objects are cards with the names/pictures of friends of the participant). The average rank of an item in a group of participants indicates how the item is perceived with regard to the attribute in relation to the other items.

The result of such a task is one-dimensional, ordinal data. It is possible to transform the resulting monadic attributes into similarity values between the participants (dyadic), indicating whether there are groups with specific views. In our case study, we use the rank order technique in one of the tasks to evaluate how “difficult” it was to make a specific piece of ceramic from the participant’s point of view. As this is an inductive study, the underlying goal was that participants were expected to apply their
own interpretations of what “difficult” meant. Similarities in placement along this rank order would then represent similar interpretations of “difficult.”

2.2. Spatial Arrangements
The Spatial Arrangement Method (SpAM) proposed by Goldstone (1994) can be seen as an extension of the rank order task. The aim of the method is to capture the similarity of items as perceived by the participants.

Given a set of items (i.e., real objects or cards with images), the participants are asked to arrange them on a square surface (or a computer screen) such that “objects that are more similar are placed closer together and objects that are less similar are placed further apart”. The question can be asked in an undirected way or with a specific criterion of similarity, e.g., “made by similar groups of people” or “serving a similar purpose.”

Once the participant is satisfied with their arrangement, a picture is taken of the whole square surface. Using an image editor, the pixel positions of the corners of the square surface and the center of each item (sherds) are tagged and recorded. (In the case of the arrangement of items on a screen, the pixel coordinates of the items can directly be stored.) Spatial Arrangements thus provide two-dimensional data on an interval scale.

Even though the method has been critiqued for having caveats in comparison to asking for a pairwise evaluation of distances (Verheyen et al. 2016), it has been shown to produce good results in a fraction of the time required for pairwise evaluation or similar methods, such as triad tests (Hout and Goldinger 2016; Hout, Goldinger and Ferguson 2013). As the field evaluators had to drive over 4,000 kilometres in a few weeks to connect with all of the study participants, time was an essential component.

The presumption of the method is that this leads participants to focus on the most important similarities, thus mimicking the intuitive construction of something similar to a principal component analysis. When averaging the results of a sufficient number of participants, the resulting distance matrix should provide a close representation of the (average) mental model of the distances of the participants. With the study that used an undefined meaning of “similarity,” these distances also allow us to assess who is interpreting (i.e., perceiving) the ceramics using similar attributions.

An obvious drawback of the SpAM method is that the placement of the objects is restricted by the two-dimensional space. As soon as more than three objects are examined, the distances between the pairs of objects can no longer be chosen independently although participants are able to move the original pair (and indeed they often did in our study).

3. Case Study Design
We presented a preselected set of 30 archaeological sherds (Figure 1) from (mostly) prehistoric contexts to a number of individuals from different groups: archaeologists with and without a specialization in ceramics, Indigenous potters, non-Indigenous traditional technology potters and two people from the general public as a control group. We asked the participants to arrange the sherds on a canvas according to different criteria (Figure 2). For the execution of the test we closely followed a written protocol in order to avoid bias induced by the formulation of the tasks (see the Supplementary Material for the protocol). In the following sections we briefly describe the tasks for the test.

3.1. Participants
We chose the participants from five different groups of people: the general public (gp), general archaeologists without specialization in US Southwest pottery (ga), ceramic analysts with a specialization in the analysis of ceramics from the US Southwest (ca), non-Indigenous traditional technology potters with European ancestry (nip),...
and Indigenous potters (ip). Subjects were chosen if they responded to a call for participants for the project. This call was generalized so as to avoid bias based on participants’ previous understandings of what we were attempting to examine. Indigenous potters were reached through the cultural divisions of their respective tribal governments. For this pilot study we interviewed:

- 6 Indigenous potters (ip)
- 4 non-Indigenous potters (nip)
- 5 ceramic analysts (ca)
- 4 general archaeologists (ga)
- 2 people from the general public (gp)

3.2. Choice of Sherds
In order to keep the tasks within a reasonable time frame to avoid participant fatigue and to provide enough space on the canvas, we decided to limit the study to a total of 30 different sherds (Figure 1). We tried to select the sherds in such a way that they provided a representative sample of the pottery found throughout the US Southwest region. While it would have been desirable, it was not possible to obtain sherds of the same size from all types we wanted to incorporate, so there is some variation in size. However this seems to not have had a large influence on the results, as only one of the participants reported size of the sherds to be one of the (minor) factors influencing their arrangement during one task. Size also does not explain clusters in any significant way.

A large portion, although in most sites not a majority, of sherds from the US Southwest are decorated. This is reflected in the choice of our sample by incorporating both painted and textured sherds. We added a few undecorated sherds of brown and gray ware. We also included one modern sherd (sherd 12) from a vessel that was made by Stella Shivwits from Acoma Pueblo. Whenever possible we chose rim sherds, as they allow a better estimate of the vessel’s shape and size. In order to limit bias based on taphonomic processes, we tried to choose sherds with as little erosion/corrosion as possible and made sure that each sherd had a fresh break to allow participants to examine the material composition and the firing process if they desired. The sherds were selected by two of the authors who are experts in Southwest archaeology (Lewis Borck and Leslie Aragon) and were photographed and recorded based on established ceramic analysis techniques for Southwest pottery. In addition, we recorded the attributes according to the Code Book for Caribbean Ceramics from Leiden University.

3.3. Tasks
After general questions regarding demographics and familiarity with pottery and archaeology, participants were asked to arrange the sherds on a 5-foot by 5-foot canvas, according to our spoken criteria. After each task, a photograph of the final layout was taken and for a few pairs of sherds (selected based on unusually large or small distances), the participants were asked why they were placed in this way. The sherds were sorted by each participant according to the following directions:

3.3.1. Warm-up Example
In order to prepare the participants for the upcoming tasks, we started with a set of 11 Lego bricks of varied shapes (different number of “pegs”, flat or tall, long or square) and colours. The participants were asked to arrange the Legos such that the Legos that are more similar to each other were closer together and those that were more different were further apart. Participants were not given any more instructions and were expected to use their own criteria to interpret what we meant when we said different. Some choose colour, some the number of pegs, some shape. While this warm-up has the slight potential to skew the participants’ later decisions for how to interpret similarity/difference, it was decided that a warm-up exercise was necessary so as not to have to dismiss the first participant task if there was an error in their understanding of the directions.

3.3.2. Task 1: Two-dimensional arrangement without guidance
For the first task participants were asked to arrange the sherds so that the sherds that were most similar to each other were closer together and those that were more different were farther apart. Here, the goal was to explore underlying perceptions of attributions based on how participants interpreted “similar” and “different”.

Figure 2: A spatial arrangement test carried out with 30 different sherds in the US-Southwest. a) Participants arrange the sherds on a canvas. b) Pixel positions of the sherds are tagged and transformed into the unit square.
3.3.3. Task 2a: Two-dimensional arrangement by perceived origin
In the first part of the second task, participants were asked to arrange the sherds so that the sherds they thought were made by a similar group of people were closer together and sherds from different groups were farther apart.

3.3.4. Task 2b: Two-dimensional arrangement by perceived function
During this task, the instructions were to place sherds from pots with a similar function closer together and sherds from pots used for different purposes farther apart.

3.3.5. Task 3: Ranking by “difficulty to make”
In the last task, the participants were asked to place the sherds on a line, ranking them from most difficult to make to easiest to make. “Difficulty” was the word that the researchers were interested in examining how different groups interpreted.

3.4. Analysis
In the following, we describe geosocial methods (i.e. joint geospatial and social networks sensu Borck 2016; Borck et al. 2015; Borck and Mills 2017; Hill et al. 2015; Leidwanger et al. 2014) that can be used to analyse the spatial arrangements and distance matrices obtained from the methods described above.

Before beginning the analysis of the arrangements, all positions have to be projected into the unit square. If the positions are recorded from a photograph, we first applied a homography to correct distortions of the arrangement induced by the angle of the camera. The necessary parameters for the transformation from the quadrilateral marked by the positions of the four corner points to the 1 × 1 square can be calculated based on the methods described by Criminisi and colleagues (1999). Then all pairwise (Euclidean) distances of the items were computed and normalized such that the largest distance between two objects was 1.

3.4.1. (Classical) Multidimensional Scaling
Multidimensional scaling can help to visualize distances between objects by projecting them onto a plane. We used this technique to visualize the average perceived distances of sherds for the groups as well as how much the answers differed between the participants.

After averaging the distance matrices from multiple participants in a two-dimensional arrangement exercise, we acquired a (symmetric) matrix of distances between n objects. In most cases, it is not possible to find an arrangement of the objects in the two-dimensional space that represents all distances.

Multidimensional scaling (MDS; Torgerson 1951, 1958) is a method to project the n-dimensional space to the k “most relevant” dimensions. The most commonly used methods for generating MDS representations use the dimensions that are spanned by the eigenvectors of the largest eigenvalues of the distance matrix. In order to create visual representations, we set k = 2, giving us two-dimensional positions of the n items that most closely represent their distances in the given distance matrix.

We use the function cmdscale from the R-package stats (R Core Team 2017) to calculate the classical MDS representations in the following sections.

3.4.2. Modularity
Modularity (Newman 2006; Newman and Girvan 2004) is a widely used concept from the field of network science to measure how well the structure of a network corresponds to a given division into clusters (“modules”). We used the concept to analyze how well the arrangements of the sherds corresponded to recorded attributes of the sherds.

Modularity is defined as the fraction of intra-cluster edges minus the expected fraction of intra-cluster edges in a random network with the same degree distribution.

For weighted graphs it is defined as follows:

$$Q = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

A_{ij} is the weight of the edge between node i and node j, k_i the weighted degree of node i (i.e., the sum of the edge weights of all edges incident to i), m is the sum of all edge weights, c_i is the cluster of node i, and \( \delta(c_i, c_j) \) is a function that is 1 if \( c_i = c_j \) and 0 otherwise.

In this case study we use the R-package igraph (Csardi and Nepusz 2006) to calculate modularity on weighted similarity matrices.

3.4.3. Mantel Test
In order to quantify how the arrangements of sherds differ between participants, we need a measure for comparing distance matrices.

A common method used to determine the correlation between two distance matrices is the Mantel Test, first introduced by Mantel (1967). Since the values in a distance matrix are not independent, a simple correlation coefficient on the values of the matrix would not produce meaningful results for the significance of the test. The Mantel test is based on a high number of random permutations of the rows and columns of the matrices where the significance is the proportion of permutations that lead to a higher coefficient.

We use the function mantel from the R-package vegan (Oksanen et al. 2010) to calculate correlations between distance matrices in the following sections.

4. Results
While most ethnoarchaeological studies have fewer participants than ours (for example 17 participating households in DeBoer and Lathrap’s excellent 1979 study on Shipibo-Conibo ceramics), we still caution that the following results are preliminary. This is because of the complicated nature of mixing ethnographic and psychological approaches. As such, we would caution against drawing any firm historical conclusions. However, the results do indicate that these methods can serve as a useful tool in quantifying and understanding cultural and institutional biases. This is one of the first steps necessary towards approaching standpoint theory’s ‘strong objectivity’ (Harding 1991, 2005).
4.1. Spatial Arrangements (Tasks 1 and 2)

We analyzed the spatial arrangements in the first three tests by averaging the normalized distances of the sherds for each group of participants. Through plotting an MDS of the average distances for each group, we constructed general patterns for the groups. See Figure 3 for an example.

In order to visualize the perceived distances for all participants and all pairs of sherds, we created a matrix of histograms of the distances. For each pair of sherds, the matrix shows a histogram of the distances between the sherds in the arrangements of the participants. Figure 4 shows the results from Test 2a (the results from the other tests can be found in the supplementary material). We placed the histograms based on the distances of all participants in the lower left half of the matrix and the ones for Indigenous potters and ceramic analysts in the upper right half, colour coded by group.

The figure shows, for example, that there was not much agreement whether sherd 2 and 7 share a common origin, as indicated by the even distribution of distances. For sherds 22 and 26 on the other hand, all participants agreed that they must have had a similar origin by placing them next to each other. The graphic also enables us to detect differences between the groups. Take sherds 19 and 16: here, all ceramic analysts agree that they have a similar origin, whereas most Indigenous potters placed them quite far apart. A reverse situation can be observed for the pair of sherds 18 and 29.

4.2. Modularity Scores

Next we quantified to what extent each of the sherd attributes is reflected in the arrangements. To do this, we transformed the distances into a network of similarities by inverting the values (i.e., the weight of an edge between two sherds is 1 divided by their normalized distance).

We calculated the weighted modularity of the clusters generated by the attributes on the resulting networks. Figure 5 highlights positive (green) and negative (red) modularity values for selected attributes on the similarity matrices based on the aggregated distances per group. The attributes (listed below) are shown in the figures and pulled from the Code Book for Caribbean Ceramics (see the Supplementary Material for a detailed description of the attributes):

- vessel shape (vs)
- wall profile (wp)
- lip shape (ls)
- rim profile (rp)
- decoration (dec)
- colour outside (co)
- colour inside (ci)
- firing atmosphere (fat)
- surface finishing outside (sfo)
- surface finishing inside (sfi)
- slip (slp)
- corrugation (cor)
- smudging (smu)

Note that the absolute scores of modularity are quite low. This is because the networks are complete networks with weights resulting from the inverted distances. These low absolute scores make modularity scores close to the (theoretical) maximum of 1 unlikely (and potentially impossible). However, the relative scores can be an indicator of how well an arrangement represents a clustering.

A positive modularity score for an attribute indicates that the average arrangement of the sherds by the participants in each group represents the clustering generated by that attribute better than expected for a random placement of the same sherds. Note that, since we are

---

Figure 3: MDS of the average distances between the sherds in the recordings of the Indigenous potters for Test 1.
analysing the average distances, the scores indicate how well the participants in the group agree on a particular clustering.

It turns out that in almost all arrangements, decoration (dec) is the primary driving factor with the exception of perceived function (Test 2b) where the non-Indigenous potters (nip) and ceramic analysts (ca) focused more on wall profile (wp) and the presence of slip (slp).

Importantly, however, the table for the undirected arrangement in Test 1 (Figure 5a) shows that even though all groups put an emphasis on decoration (dec) in their arrangements, the strengths of the agreement varies. While general archaeologists’ (ga) modularity is the highest of all tests and groups, Indigenous potters (ip) seem to put (on average) a lot less emphasis on the decorated criterion. This is to say, it is important, but less so than with other groups.

Interestingly, the table for the arrangement based on the perceived origin in Test 2a (Figure 5b) shows a reverse picture: Here, the Indigenous potters (ip) based their average judgment more on decoration than the general archaeologists (ga).

In Test 2b, asking for an arrangement by function (Figure 5c), general archaeologists (ga) and Indigenous potters (ip), who diverged in previous tests, have similar emphasis in their average arrangements and rely on decoration (dec) as an important factor. Non-Indigenous potters (nip) seem to agree more with the ceramic analysts as they put their highest emphasis on wall profile (wp) and slip (slp).
In the Supplementary Material we provide additional tables of the modularity for each individual arrangement.

4.3. Comparing Individual Arrangements

In order to compare the individual arrangements of the sherds, we calculated the distances between the recorded distances of the participants. For each pair of participants, we calculated the Euclidean distance between the vector of the 435 normalized pairwise distances between the sherds. Figure 6 shows the MDS from the resulting distances for Test 2a. Participants are color-coded by the group they belong to. The image suggests that there is a higher diversity within the group of Indigenous potters (IP) than among the other groups. The group of ceramic analysts (CA) is placed at the closest distances to each other, suggesting a higher homogeneity in their arrangements.

A possible explanation for this could be that ceramic analysts have formal training on the subject and therefore agree more consistently because of this institutional background. This is why they cluster in the centre. Indigenous potters in contrast apply different criteria/conceptualizations, resulting in larger differences between their arrangements.

**Figure 5:** Modularity of the clusterings induced by selected attributes on the inverted distance matrices. **a)** Test 1: unguided arrangement. **b)** Test 2a: perceived origin. **c)** Test 2b: perceived function.

**Figure 6:** MDS of the distances of distances for the participants for Test 2a.
4.4. Ranking of the Sherds (Task 3)

The boxplot in Figure 7 shows the variability of the rankings for the different sherds ordered by the median value. The higher the ranking the more difficult it was to make the sherd according to the participants. While there is a relatively strong agreement about the sherds that were easiest to make (the undecorated sherds 1, 3, and 7), opinions diverge more when it comes to the sherds whose median ranks are between 10 and 20. The highest variation in the answers was observed for sherd 12, which is the strikingly bright-white sherd from the modern pottery vessel made by Stella Shivwits (Acoma Pueblo). While considered “hard to make” by some of the participants, others rejected the sherd as being “fake” (yet not necessarily badly made). Other sherds for which opinions diverged include sherd 2 (which is very thin and fired well, but not decorated), sherd 20 and 28 (both are corrugated), and sherd 8 and 9 (they are both polychrome).

The overlay of the scatter plot showing the individual rankings colour coded by groups hints at an interesting pattern: Almost all outliers of the top-ranked sherds as well as the lowest-ranked sherds come from participants within the group of Indigenous potters (ip).

In order to examine this trend more closely, we established distances between the participants based on their rank correlation. We calculated Kendall’s $\tau$ on the rankings and generated a distance matrix based on the inverted values (i.e., $1-\tau$). Figure 8 shows an MDS of the inverted

![Figure 7: Boxplots of the participant’s rankings of the sherds in Test 3, ordered by mean value, overlaid by a scatterplot of the individual ranks coloured by group.](image1)

![Figure 8: MDS of the inverted rank correlation calculated with Kendall’s $\tau$ for Test 4.](image2)
correlations. Clearly the Indigenous potters are spread out much more than the rest of the groups. This likely indicates that within this group there is a higher diversity in the interpretation of what it means that something is “difficult to make.”

4.5. Outlook

While the sample size in our study is not large enough to make statistically reliable statements about the differences in perception between different groups, the results indicate that such differences likely exist and are measurable with the methods we propose. If the results are to be used to analyze “folk” classifications versus Western scientific classifications, the number of participants would need to be greatly increased. Regardless of which particular avenue of interest is followed, we believe that the results of this study are promising enough to extend the survey. The methods presented here could also provide valuable insights in other areas of anthropology where triad tests might prove too cumbersome and time intensive for participants.

5. Conclusion/Discussion

While this was an exploratory project, the differences and similarities found between groups highlight some important areas for discussion. For instance, when interpreting the term “difficult” from Test 3 both groups of archaeologists frequently focused on technique in painting and, more rarely, skill at firing. Thus, across both archaeological groups, difficulty was interpreted as skill, almost exclusively in either stylistic implementation and object manufacture. Non-Indigenous traditional technology potters often followed very similar interpretations of language as both archaeological groups and often sorted in similar ways. This hints at a similarity in underlying cognitive processes in constructing taxonomies. This may be a product of expertise arising, often, from Western science based perspectives. This is perhaps unsurprising as most of the non-Indigenous traditional technology potters are well-versed in the archaeological literature.

Indigenous potters were the least likely group to sort by decoration. Their lower levels across many of these decorative variables indicate they are relying on more variables. Thus, Indigenous traditional technology potters, based on how they interpreted our key word in each task, perceived the ceramic artefacts in a much more holistic way. During follow-up interviews, and discussions with the participants about why they chose to sort in particular ways, Indigenous potters were also more likely to interpret the intentionally vague word included in each task in a more relative manner than the other participants. For instance, the actual act of pottery production is not necessarily an economic or artistic enterprise (although in the contemporary world production for the art market is definitely a part of the craft [Hoerig 2003]). As discussed by Lee Ann Cheromiah (LAC), a traditional potter from Laguna Pueblo:

“In our culture, my mother has always said you pray to the spirits, and I always say help me mom, this isn’t working. You have to connect with your clay and if you’re not connecting with your clay your mind’s not thinking about it. … I can make however many pots in just a few hours. But this one I wasn’t in the right frame of mind and it was three, four hours and I couldn’t build it. And my mother says, “You’re not in sync, you’re not in connection with your clay, so put it away.”

Beyond being an outside spirit, the clay also retains powerful connections to humanity. LAC continues:

“So its like this spiritual connection that we always have. And she taught us also that the clay is spiritual. And it’s a she. The clay is a she.”

Thus Indigenous potters express a more spiritual interconnection between themselves and the material comprising the ceramics. This is also partially reinforced because the production process for many of the Indigenous potters in the study also incorporated aspects of memory making through the reinforcement of ancestral connections. Pottery-making can thus create, and combine, history and emotion. As LAC explains:

“So I was getting ready for the Indian market, and I was still painting a pot. And I kept thinking, I know this pot, I can visualize the design, but I can’t see details of it. And I went over to her house [their mom had passed away a few years before] and my sister says, “Here’s a picture of a design mom was working on”. And I walked up and took the sketch and I said, “This is the design I’ve been searching for!” And then I said, “Well I came over after some paint you know. I need some paint. Real paint.” And she said okay and she gave me the paint. And so I came back up here (their own home) to soak my real paint (red clay) and … this paint it smells so good. The clay it has a particular scent to it. What is it? And then I’m like, “Ahhhhhhhh, it’s my mother.”

Indigenous potters were also more likely to interpret the intentionally vague word included in each task in a more relative manner than the other groups. For example, one potter noted that a sherd that most participants had discussed as belonging to a ceramic vessel that would have been easy to make, would instead have been difficult to make. This was not an evidence based error either. The potter noted that there were some markers on the sherd
that indicated to them that the vessel the sherd was originally a part of was made by a beginner. Thus, the pot was difficult for a beginner to make. In this instance then, difficulty was interpreted relative to the skill of the ancient potter. For this Indigenous potter, at least, difficulty depended entirely on the ancient potter. They did not assume the proficiency of the ancient potter.

Another notable difference between the four groups may relate to diverging ideas on what traits are more closely linked with ceramic vessel functionality (Test 2b). For example, the non-Indigenous potters and the archaeological ceramic analysts sorted with a strong focus on vessel shape, wall profile, and slip. This was not something that the general public or the Indigenous potters focused on. Moreover, the presence of decoration was not driving the non-Indigenous potters’ vessel function spatial arrangements. Decoration strongly drove the Indigenous potters’ and the general archaeologists spatial arrangements though. These differences may indicate underlying ideas about functionality that focus on storage versus serving tasks for ceramic vessels. This was a primary point brought up in follow-up interviews, particularly with ceramic analysts. Indigenous potters’ focus on decoration and slip as a driving factor for spatial sorting when asked about the function of the ceramic vessels that the sherds were from might indicate that (as many Indigenous potter participants noted during follow-up discussions) they saw decorations, and not form, as being related to ceremonial functions of the artefacts.

Notably, two tests (2a – perceived origin and 3 – difficulty to make) highlight arguments proposed by Feyerabend (1993), who argued there is problematic orthodoxy and dogma within science, and supporters of integrating Indigenous science with Western modern science (WMS; e.g., Cajete 1999; Corsiglia and Snively 2001; Snively and Corsiglia 2001) who have noted that Indigenous models often emerge from dramatically different foundations than WMS. For example, in Test 2a (e.g. Figure 6) it is probable that archaeological ceramic analysts are clustering near the center because they are using similar evaluations that they have gained through the construction of orthodox expertise in the discipline. General archaeologists and non-Indigenous potters show more variation, and thus less orthodox. Indigenous potters, on the other hand, are not clustering, likely because they are applying different criteria and different conceptualizations gained through non-WMS avenues for expertise construction. Essentially, their expertise is not institutionally homogenized.

Again, while this study sample is not large enough to draw clear conclusions, the continual placement of Indigenous traditional technology potters outside of the median does at minimum suggest that the current archaeological mode of creating typologies and analyzing ceramics needs to be re-evaluated to incorporate how Indigenous potters perceive and interpret ceramics if the discipline is concerned with knowledge systems not well represented by Western modern science. There is no “one-size-fits-all” approach here, either. When Indigenous potters were outliers, they were often outliers along with other members of their community and not with Indigenous potters from other communities. This should be unsurprising given that Indigenous groups have diverse cultural foundations through which they perceive the world. This reflects a need for archaeologists to explore new, and flexible, ways to examine ceramics that can incorporate culturally diverse perspectives while maintaining analytical comparability for the archaeological discipline. In many ways, this highlights that we need to implement what some (e.g., Caraher 2016; Cunningham and MacEachern 2016) have called slow archaeology, and what Wang (2013) has called “thick data,” to better contextualize our big datasets.

As with all ethnoarchaeological approaches, the contemporaneity of the individuals that were participating leads to concerns as to how closely these results may be representative of past communities (e.g., Cunningham 2003). If the contemporaneity of the participants was a primary factor, instead of, for example, culturally derived differences in how ceramics or material are viewed, then our tasks should display either a broader spread of placement in relationship to variables and/or more consistent outliers from the other non-archaeologist groups. Instead, this may indicate that, as Cunningham has noted (2003), ethnoarchaeological projects, instead of simply being middle range theory building for archaeological analogy, could be used to complement and support Indigenous forms of scholarship on the material past.

Once again, we want to stress that this is an exploratory analysis. Future studies should build on this by increasing sample sizes to effectively explore these potential differences. We also need to incorporate sherds from other archaeological regions, as well as local and researcher participants living and working in those contemporary regions to see if patterns and differentiations are present in other areas between groups as well. An eventual output could be to start to explore how adjusting the weights of various ceramic attributes within typologies may help to create taxonomies that are useful for interpreting culturally relative processes embedded within ceramic material.

Additional Files

The additional files for this article can be found as follows:

- Supplementary Material 1. Appendix. DOI: https://doi.org/10.5334/jcaa.37.s1
- Supplementary Material 2. R-scripts. DOI: https://doi.org/10.5334/jcaa.37.s2
- Supplementary Material 3. Data. DOI: https://doi.org/10.5334/jcaa.37.s3

Acknowledgements

First, we want to acknowledge all of the participants in this study. Your conversations with us and the time you took out of your day to play with ceramics with us was highly appreciated. This research has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 319209 and the Netherlands Organisation for Scientific Research (NWO) project number 360-62-060. Human subject research was reviewed and approved by the Faculty of Archaeology at Leiden.
University research committee and through Unit B1, Ethics Review and Expert Management with the European Research Council. Ceramics were loaned to us by Barbara Mills and the Southwest Laboratory at the University of Arizona and Archaeology Southwest.

Competing Interests
The authors have no competing interests to declare.

References


Binford, SR and Binford, LR. 1969. Stone Tools and Human Behavior. Scientific American, 220: 70–84. DOI: https://doi.org/10.1038/scientificamerican0469.70


How to cite this article: Borck, L, Athenstädt, JC, Cheromiah, LA, Aragon, LD, Brandes, U and Hofman, CL. 2020. Plainware and Polychrome: Quantifying Perceptual Differences in Ceramic Classification Between Diverse Groups to Further a Strong Objectivity. *Journal of Computer Applications in Archaeology*, 3(1), pp. 135–150. DOI: https://doi.org/10.5334/jcaa.37

Submitted: 13 March 2019 Accepted: 19 January 2020 Published: 23 April 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

*Journal of Computer Applications in Archaeology* is a peer-reviewed open access journal published by Ubiquity Press.