4

A Probabilistic Analysis of Ceramic Technology

Putting knowledge transmission at the heart of archaeological studies in ceramic technology requires a method which compares variability in ceramic *chaînes opératoires* (see Ch. 3). This chapter is an outline of the method I have developed for this purpose. The basis of the method is a conceptualisation of the ceramic *chaîne opératoire* as a network (Section 4.1), and a re-casting of this network as a probability space (Section 4.2). This creates a probabilistic method which considers the *chaîne opératoire* of a single vessel as one of many alternate paths through a network, and compares these alternatives in an inferential, quantitative manner. The final section of this Chapter (Section 4.3) is a comparison of this probabilistic approach to the qualitative, abductive approach commonly found in studies of ceramic technology.

4.1 The *Chaîne Opératoire* as a Network

The *chaîne opératoire* approach is crucial for tracking past knowledge in this dissertation (see Section 3.2). This section starts with an introduction to this approach, followed by a re-conceptualisation of it through network analysis, with specific attention to the concept syntax.

The term *chaîne opératoire* derives from Leroi-Gourhan (1964, 1965), but some debate exists regarding the history of the approach (cf. Audouze *et al.* 2017; Delage 2017a, 2017b). Initial applications of the *chaîne opératoire* are best known from analysis of lithic technology (cf. Creswell 1972; Pelegrin *et al.* 1988; Tixier 1967), but the approach is currently commonplace in archaeological analyses of various materials (e.g. Jørgensen *et al.* 2018; Kuijpers 2018; Miller 2009). Key works for the ceramic *chaîne opératoire* are Balfet (1965), Rye (1981), and Roux (2017, 2019a). The vocabulary from the latter publication by Roux (2019a) is adopted here to describe ceramic technology.

Leroi-Gourhan (1964 p. 164, 1965 pp. 132–3) envisions the *chaîne opératoire* as a dialogue between tools, gestures, and materials (cf. Lemonnier 1992 p. 1; Mauss 1936 for the definition of technique). This dialogue is ordered into a logical sequence from raw material to finished product by a syntax. The dialogue can be read from traces left on manufactured items as a sequence of techniques (cf. Roux 2019a for ceramics). This definition of the *chaîne opératoire* as a sequence of techniques which obeys a syntax is crucial for the definition of the network below.

The *chaîne opératoire* approach is not universally accepted. Ingold (2010, 2013) criticises the approach for leading to a conceptualisation of production processes as rigid, linear sequences (cf. Sofaer 2018 for ceramics). As shown below, the network representation can overcome this critique by showing this critical conceptualisation and the classic notion of the *chaîne opératoire* outlined above are two sides of the same coin.

The sections below are a re-conceptualisation of the *chaîne opératoire* through network analysis. Network visualisations of the (ceramic) *chaîne opératoire* are common (e.g. Gosselain 2018 Fig. 1; Lemonnier 1992; Miller 2009 p. 108; Roux 2019a Fig. 2.42), but these networks seldom serve as analytical tools (cf. Brysbaert *et al.* 2012; Kuijpers 2018). In this respect, the network approach developed below pushes beyond the existing use of networks in studies of past technology.

Defining the Basic Elements of the Network

A network is a mathematical abstraction which shows the structure of relationships between entities. The abstraction presents the entities as nodes, and their relations as edges (cf. Newman 2010 p. 1). The network representation developed here takes the inventory of techniques defined by Roux (2019a; e.g. coiling, scraping, and burnishing) as nodes. An edge between two techniques in the network indicates a technique can follow up another technique. For example, the network below has nodes for the roughing-out technique modelling and the preforming technique scraping. An edge joins these nodes, because a potter can follow up a modelling operation by scraping the surface of the rough-out. I argue below these edges are directional, but first two brief remarks on the choice of nodes in the network.

Firstly, the equation of nodes to techniques is not a necessary feature. Provided the arguments below regarding the directional edges in the network holds, the same analysis can be applied at coarser resolutions (e.g. stages of the *chaîne opératoire*, such as roughing out, preforming, and surface treatment) and at finer resolutions, such as modalities of techniques like coiling by spreading or coiling by pinching (cf. Roux 2019a pp. 41–2 for definitions).

The choice to employ techniques rather than stages or modalities of techniques is a balancing act. For reasons outlined below, the optimal data for the probabilistic comparison consist of complete *chaînes opératoires* from start to end. Obtaining such complete *chaînes opératoires* is nearly always possible for archaeological ceramics at the level of stages, which implies all ceramics can be included in the analysis (high representativity). However, the outcomes would be generic (low specificity), because the sequence of stages is nearly always the same in ceramic *chaînes opératoires*, regardless of the context. Vice versa, obtaining complete *chaînes opératoires* at the level of modalities of techniques would enable highly specific results at the cost of low representativity, as few complete *chaînes opératoires* can typically be reconstructed at this level of resolution from fragmented archaeological assemblages. In other words, the higher the resolution of the nodes, the more specific the outcomes but the lower the representativity due to the smaller amount of data, and vice versa. As such, performing the network analysis at the level of techniques is not a necessity, but offers a practical balance between specificity and representativity.

The second remark is about the absence of nodes for specific raw materials and paste preparation processes in the networks below. Similar to the choice to employ techniques as nodes, the absence of these two stages is not a necessity, but a practical choice relating to two factors.

Firstly, there is a difference in detection methods and sampling strategy (see Ch. 5). This study uses ceramic petrography to study raw materials and paste preparation, and relies mostly on macroscopy with some input from ceramic petrography for the detection of other production techniques. Contrary to macroscopy, ceramic petrography is a destructive method, and not all vessels studied through macroscopy could also be sampled for thin sections. Consequently, the two datasets differ in size and composition (see Ch. 5).

The second issue relates to classification and is the chief problem. Ceramic petrography revolves around geological classifications of sediments and rocks. The relation between these classifications and emic classifications of raw materials is not straightforward (cf. Arnold 1971, 2018). The construction of perceptive categories for raw materials could overcome this issue (cf. Kuijpers 2018), but such an undertaking is beyond the scope of the present study. The matter lies differently for pottery production techniques which are well-attested in ethnographic and archaeological studies (cf. Roux 2019a).

Given the issues with classification and differences in samples, correlation between the datasets on raw materials use and preparation and the dataset on techniques is more prudent than direct integration through probabilistic analysis. Hence the absence of nodes for raw materials choice and preparation in the networks below.

Following these two remarks on the definition of nodes, let us look at the nature of the edges in the network. The key point in the next section is that the edges between nodes are directional because of the syntax of the ceramic *chaîne opératoire*. Directionality is a crucial element for the probabilistic analysis, and enables incorporation of complex patterning in ceramic technology.

From Syntax to Directional Network

Syntax is a key element in the definition of the *chaîne opératoire*, but plays a minor role in applications and comparisons. This section presents an argument for a formal connection between syntax and hydric states in ceramic production. Network analysis can incorporate and compare this syntax by making the edges between nodes directional.

Leroi-Gourhan (1964, 1965) uses the term syntax as a metaphor for the logic which binds tools, gestures, and materials into a coherent sequence. Lemonnier (1980 p. 9, 1992 pp. 21–2) elaborates on the concept syntax by proposing *chaînes opératoires* consist of strategic and flexible steps. Strategic steps cannot be postponed, cancelled, or reversed, because they are crucial elements within the syntax of a production process. Flexible steps on the other hand can occur at any time, or be ignored, altered, and reversed. Following this definition, Gosselain (2018) identifies acquisition and processing of clay, shaping, drying, and firing as strategic parts of the ceramic *chaîne opératoire*. The discussion below picks up on the connection between the latter two items on this list. We return to raw materials and shaping at a later stage. For now, let us say pottery production self-evidently requires clay and temper (raw materials), as well as roughing-out techniques which transform these raw materials into a vessel.

Gosselain (2018) states drying and firing are strategic steps in the ceramic *chaîne opératoire*. Both steps relate to hygrometry, which is a measure of the humidity of a body (cf. Roux 2019a p. 43). The hygrometry of a clay body affects its mechanical properties. For our purposes, the chief mechanical property affected is plasticity: the capacity for permanent deformation under applied force. A clay body with high humidity is plastic and can be shaped by hand, whereas shaping by hand is impossible for a clay body rendered rigid and brittle by low humidity. Consequently, the application of the various ceramic production techniques depends on the hygrometry of the clay body, because of this relation between hygrometry and plasticity (cf. Roux 2019a p. 43).

The relation between hygrometry and the application of techniques is a key element for the directionality of the network. Hygrometry is a continuous variable: The clay paste dries up gradually during the production process as a result of deliberate drying or ambient temperatures. Moreover, a successful ceramic production process always develops from high humidity to low humidity (cf. Roux 2019a p. 43) even if this does not necessarily imply time pressure (cf. Sofaer 2018). The directional nature of the production process necessarily follows from this continuous nature of hygrometry: any two techniques form a sequence because the clay ever so slightly dries as the production process goes on, and because the application of techniques depends on the hydric state of the clay body. For example, a potter can go from coiling to burnishing, because burnishing requires a drier clay body than coiling. However, the inverse order is impossible because that would require the clay body to become moister over time.

Sofaer (2018) argues hygrometry is not irreversible in a strict sense. Potters may opt to soak and dissolve a dry vessel, or re-humidify a vessel surface during production. These observations are valid, but over-stated. Re-humidification of vessel surfaces, for example during softening, does result in traces which are distinct from those left by operations on wet clay (cf. Lepère 2014; Roux 2019a pp. 200–1). Moreover, dissolving a dry vessel into wet clay and re-doing the production process implies erasing all traces of the techniques preceding the dissolution. Therefore, any finished vessel will exhibit a directional development from high to low humidity regardless of such operations. As such, the directionality of ceramic production processes holds.

Despite the continuous nature of hygrometry, archaeologists generally distinguish four discrete hydric states for clay bodies (Roux 2019a p. 43). From high to low humidity the hydric states are: wet, leather-hard, dry, and fired. Crucially, ceramic production processes always pass hydric states in the above order. None of the hydric states can be skipped: Firing a leather-hard vessel would result in failure (Roux 2019a pp. 43; 110). The (sequence of) hydric states in ceramic production should not be considered the product of modern measurement techniques. The fact that prehistoric potters produced vessels automatically implies their awareness of the crucial role of hygrometry in ceramic production. These hydric states can probably be precisely replicated by means of perceptive categories (sensu Kuijpers 2018) given the differences hygrometry induces in surface appearance, malleability, and texture of a clay body (cf. Rye 1981 pp. 20–1; 24).

The above paragraphs argue hygrometry *is* the syntax of the ceramic *chaîne opératoire*. The relation between hygrometry and plasticity of a clay body implies hygrometry orders the application of techniques during the production process, because techniques can only be applied to clays with certain degrees of plasticity. Moreover, the directional development from high to low humidity in the ceramic production process can be influenced by potters, but not altered, postponed, cancelled, or reversed in a successful production process.

The network representation of the ceramic *chaîne opératoire* captures the syntax of the production process by making the edges between nodes directional. Two techniques form a sequence because the clay is ever so slightly drier after the application of each technique. The directionality of the edges reflects this continuous, progressive change in the hygrometry of the clay body. The network acknowledges the influence of potters over hygrometry by also including nodes for drying to leather-hard and dry consistency, as well as firing of vessels. This puts the drying of clay vessels on equal footing with other deliberate operations, such as coiling and painting. The first node in the network is 'wet clay', in which the prepared clay paste has the highest hygrometry. This node necessarily leads onto roughing-out techniques from where a multitude of alternate sequences become possible. Hence the designation of roughing-out techniques as strategic by Gosselain (2018). Roughing-out techniques are simply the first stop on the route.

Given the relation between hygrometry and plasticity, each technique in the network also exhibits a 'tolerance' which indicates the hydric states in which application of said technique is feasible. For example, the tolerance of modelling is the hydric state wet, whereas the tolerance for burnishing are the hydric states leather-hard or dry. The information about the tolerance of techniques is based on the overview in Roux (2019a).

The discussion on the tolerance of nodes completes the definition of the network conceptualisation of the *chaîne opératoire*. The resulting network with techniques as nodes, directional edges to capture the syntax, and with specific tolerances attached to each technique is presented in Fig. 4.1. This network representation allows for the distinction of two key terms for understanding and comparing *chaînes opératoires*: total and specific *chaînes opératoires*.

Total and Specific *Chaînes Opératoires*

In this section, I propose a terminological nuance which paves the way for the probabilistic analysis and addresses recent critiques of the *chaîne opératoire* approach.

Present use of the term *chaîne opératoire* can both refer to the abstract whole of ceramic technology, and to a singular production process reconstructed from traces on a specific vessel. The network representation above enables a distinction between these two conceptualisations.

The abstract whole of ceramic technology is referred to as the *total chaîne opératoire*. The total *chaîne opératoire* consists all possible ceramic production techniques, and, crucial for the probabilistic approach below, all possible combinations of these production techniques allowed within the syntax of the *chaîne opératoire*. The network in Figure 4.1 represents this total *chaîne opératoire* for ceramic technology.

By contrast, the production process of a vessel which one can reconstruct from technical traces on the vessel is a *specific chaîne opératoire*. Given that a specific *chaîne opératoire* is a sequence of techniques, and that the total *chaîne opératoire* is the overview of all possible combinations of techniques, all possible specific *chaînes opératoires* exist as paths through the network in Figure 4.1. In graph theory, a path is a list of nodes in which each node on the list exhibits an incoming link from the previous node (Newman 2010 p. 136). The relation between network and path is a direct analogy for the relation between total and specific *chaînes opératoires*.

The relation between path and network is the cornerstone of the probabilistic approach developed in the next section. In addition, the distinction between specific and total *chaîne opératoire* reconciliates recent criticisms of the *chaîne opératoire* approach as overly rigid and linear (cf. Ingold 2010, 2013; Sofaer 2018) with classic conceptualisation of the concept (cf. Lemonnier 1992). The classic conceptualisation of the *chaîne opératoire* as a

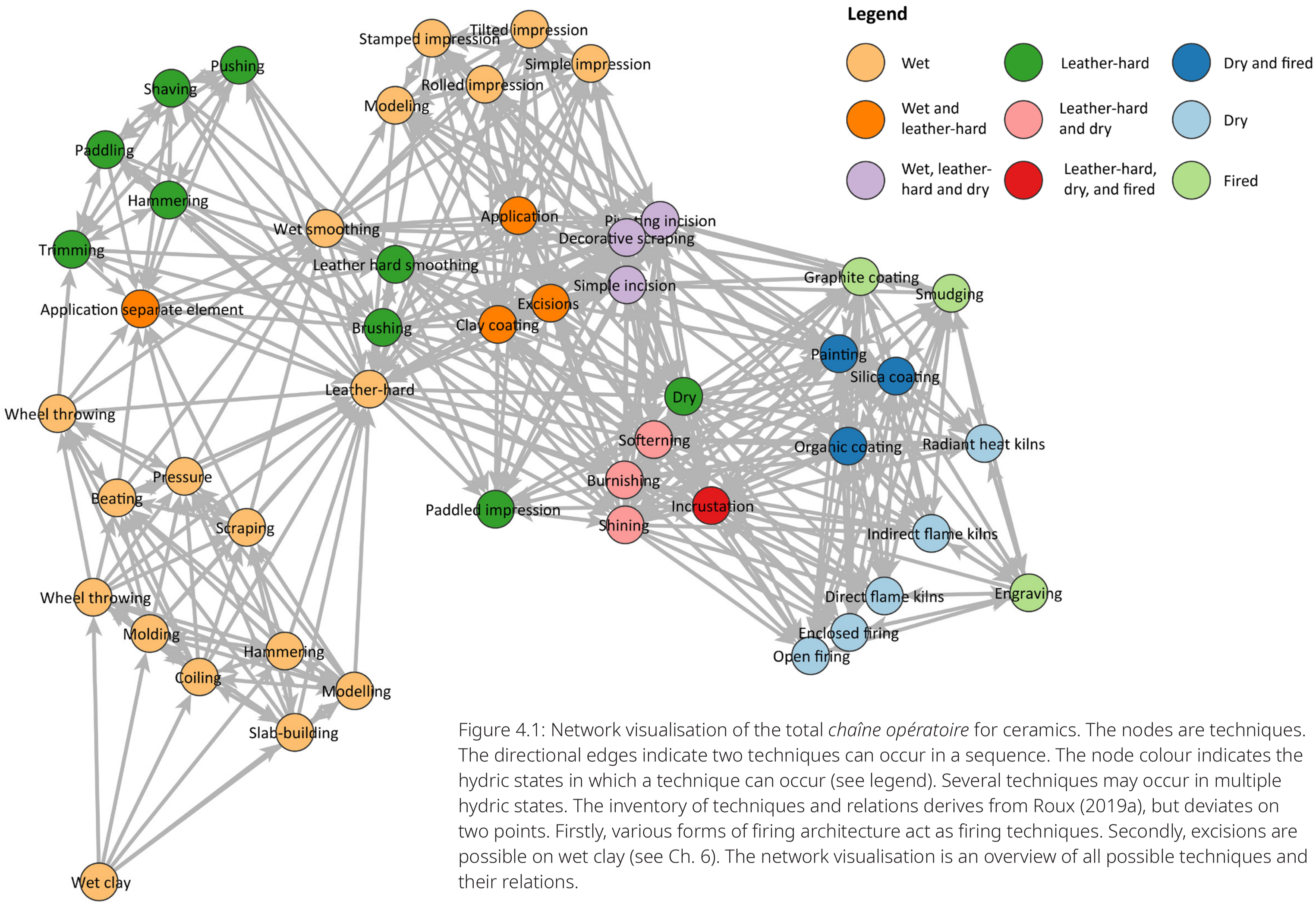

Figure 4.1: Network visualisation of the total *chaîne opératoire* for ceramics. The nodes are techniques. The directional edges indicate two techniques can occur in a sequence. The node colour indicates the hydric states in which a technique can occur (see legend). Several techniques may occur in multiple hydric states. The inventory of techniques and relations derives from Roux (2019a), but deviates on two points. Firstly, various forms of firing architecture act as firing techniques. Secondly, excisions are possible on wet clay (see Ch. 6). The network visualisation is an overview of all possible techniques and their relations.

sequence of techniques is about specific *chaînes opératoires* or paths through the network. However, the above-mentioned critics look at the total *chaîne opératoire*, the network as a whole, which indeed offers the possibility of many alternate, cyclical, and varying specific *chaînes opératoires*. Hygrometry connects these two different views: the directionality of the edges which results from hygrometry, from the behaviour of clay, implies all specific *chaînes opératoires* ultimately appear as linear sequences from high to low hygrometry through the more complex total *chaîne opératoire*. As such, the network representation resolves the contradiction between the classic and critical conceptualisations of the *chaîne opératoire* approach.

Re-conceptualising the *chaîne opératoire* also opens up possibilities for new approaches to ceramic technology. The next section is an outline of such a new approach with the network in Figure 4.1 as the point of departure.

4.2 The Total *Chaîne Opératoire* as a Probability Space

The previous section is a re-conceptualisation of the *chaîne opératoire* approach through network analysis. A key feature of the network approach is that any specific *chaîne opératoire* exists as one among many possible paths in the total *chaîne opératoire*. This insight allows for the formulation of a probabilistic comparison of specific *chaînes opératoires*. The first step in this formulation is to show the total *chaîne opératoire* can be thought of as a probability space.

Fundaments of a Probabilistic Conceptualisation

The term probability space derives from Kolmogorov (2018) and is a formal model for chance processes. A probability space consists of three elements: a sample space, an event space, and a probability function. The sample space is an overview of all possible outcomes of a given chance process, whereas the event space is a set of specific outcomes. The last element, the probability function, assigns a probability to each outcome (Chow and Teicher 1988 pp. 19–20; Sazanov 2002). The network conceptualisation of the *chaîne opératoire* above can be understood as a probability space.

The total *chaîne opératoire* is a direct analogue for a sample space. The network representation shows all possible outcomes of the ceramic production process in the form of paths through the network (see the discussion about sound, valid, and invalid paths below). A specific *chaîne opératoire* is a direct analogue for an event: one specific outcome within the sample space. Returning to the definitions laid down in Chapter 3, a body of knowledge, or any set of specific *chaînes opératoires*, can be thought of as an event space.

The crux of the approach is the definition of the probability function. No direct analogue for this element exists in ceramic technology. Instead, the approach below uses relative frequency distributions of specific sequences of techniques as a probability function (see Fig. 4.2). It works as follows. We take a body of knowledge for (f.e.) Funnel Beaker West. Every specific *chaîne opératoire* in that body of knowledge consists of a number of choices to follow up one technique with another. The total *chaîne opératoire* is an overview of all possible choices, presented as edges between nodes (see Section 4.1). We give each edge in the network a value which is equal to the percentage of specific *chaînes opératoires* from the Funnel Beaker West body of knowledge which features that choice. For example, if 90% of these specific *chaînes opératoires* saw the application of simple impressions during decoration followed up by drying the vessel to a leather-hard consistency, but the other 10% first underwent simple incisions after the simple impressions, then the edge between the nodes 'simple impression' and 'leather-hard' receives a value of 90, and the edge between the nodes 'simple impression' and 'simple incision' a value of 10. We interpret these numbers as an indication of the procedures with which potters were familiar, and the commonness of these procedures. As such, they become the probability function which assigns a likelihood to a particular choice (in the example above, simple impressions are more likely to be followed by drying to leather-hard than by simple incisions). Doing this systematically for all choices in the specific *chaînes opératoires* results in a network which captures the variability in Funnel Beaker West ceramic technology in a quantitative fashion. The most common combinations of techniques become the busiest paths through the network, while alternate options or exceptional choices branch off or form local detours. The power of this analogy with a probability space is the ability to formally compare this variability between bodies of knowledge.

A Method for Probabilistic Comparison

Building on the analogies above, a probabilistic comparison between bodies of knowledge becomes possible. This procedure consists of three steps.

The first step plots all specific *chaînes opératoires* in a given body of knowledge, say body of knowledge A, as paths through the network in Fig. 4.1. The comparison then extracts all edge weights in the network as a probability distribution (see Fig. 4.2). Relating back to Section 4.1, the probability distribution not only captures the occurrence of specific techniques (i.e. nodes) but also the syntax of the production process because it extracts all combinations of techniques (i.e. edges) in the network simultaneously (see Fig. 4.2). By repeating step 1 for a second body of knowledge (body of knowledge B) we get two such probability distributions.

The second step consists of a formal comparison between the probability distributions for body of knowledge A and B. The Wasserstein distance is employed for this comparison (see SciPy documentation 2023 for the algorithm). The Wasserstein distance is a non-parametric method to compare two probability distributions which exist in the same metric space. The comparison computes the minimal amount of work required to transform one distribution into the other (Ramdas *et al.* 2017; Villani 2009 pp. 93–4 for the exact definition). The more dissimilar the two distributions are, the more work is necessary to complete the transformation, and therefore the larger the Wasserstein distance between them (and vice versa).

In the context of comparing specific *chaînes opératoires*, the Wasserstein distance informs us about the minimum number of alternate choices potters would need to make, or new techniques they would need to learn, in body of knowledge A (f.e. Funnel Beaker West) to arrive at the specific *chaînes opératoires* observed in body of knowledge B (f.e. Corded Ware). If Funnel Beaker West and Corded Ware specific *chaînes opératoires* share many sequences of techniques, for instance, if only one surface treatment technique differs, then the comparison returns a small Wasserstein distance. The distance increases if frequently observed sequences of techniques in the Funnel Beaker West body of knowledge are marginal, or do not appear, in Corded Ware. As such, we can think of the Wasserstein distance as the odds that alternate choices within the Funnel Beaker West body of knowledge would produce the specific *chaînes opératoires* in the Corded Ware body of knowledge. The greater the distance, the more changes in the choices

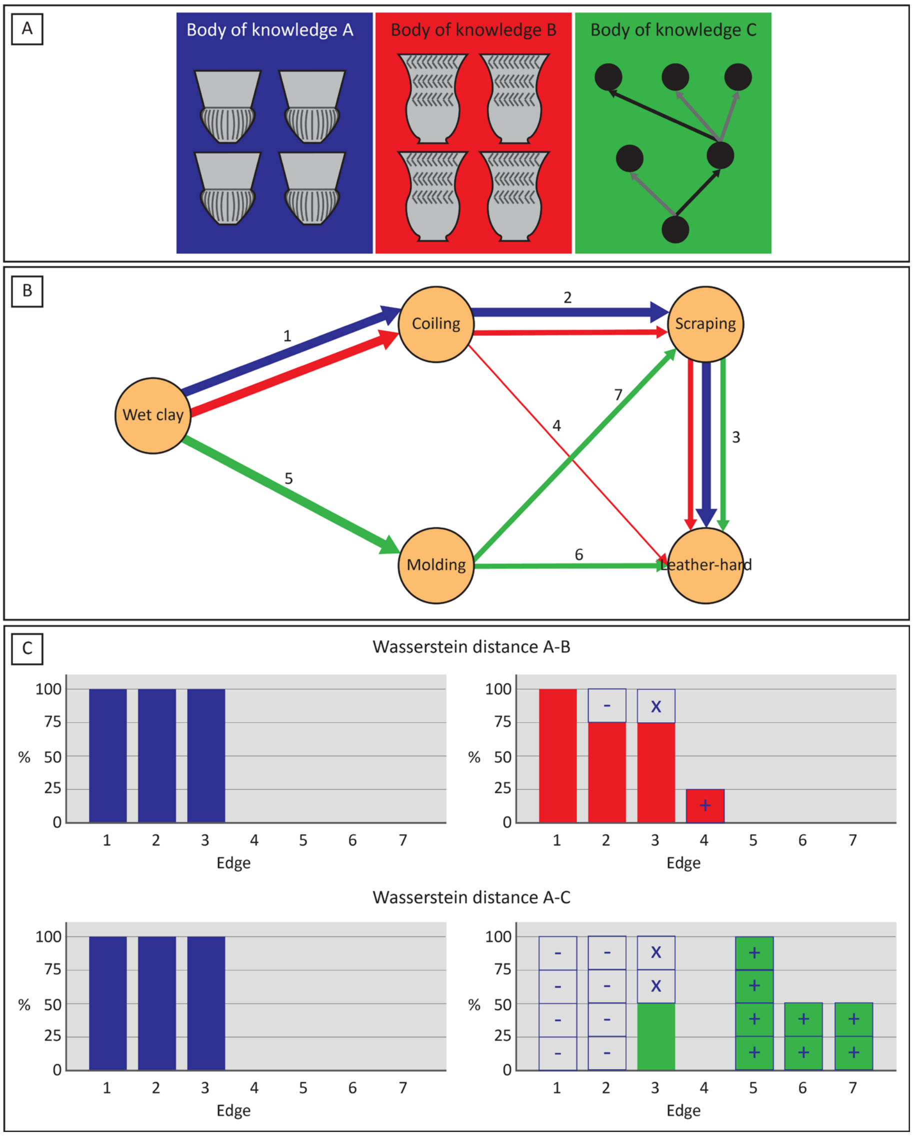

Figure 4.2: Method for probabilistic comparisons of ceramic *chaînes opératoires* in three steps (A-C). A: The method requires three bodies of knowledge (A, B, and C) at minimum. It is unknown whether the potters behind bodies of knowledge A and B learned from each other, but the potters who made body of knowledge C must not share knowledge with those behind body of knowledge A. B: The comparison represents the specific *chaînes opératoires* within these bodies of knowledge (see colours) as paths through the total *chaîne opératoire*. The numbers in the graph are labels for the edges. The thickness of the edges indicates the relative number of specific *chaînes opératoires* which run along the edge. C: The edge weights in the entire network then become a probability distribution. The Wasserstein distance calculates the minimal effort required to transform the probability distribution of body of knowledge A into those for body of knowledge B and C. Effort, in this context, means making alternate choices and learning new techniques. In this example, it requires less effort to transform body of knowledge A into B, than to transform it into an unrelated body of knowledge C. As a result, we may conclude the technical variability in bodies of knowledge A and B likely results from potters sharing a certain amount of technical knowledge (see Tab. 4.1).

and knowledge of potters required, and the less likely these potters learned the same production processes. Crucially, the Wasserstein distance is never infinite: there will be a distance, even if two bodies of knowledge do not feature even one shared technique, but this distance will simply be large. As such, the probabilistic approach always allows for the possibility, however remote, that potters learn and integrate completely new techniques (see Section 3.2).

This sensitivity of the Wasserstein distance to differences in edge weights is also why working with complete specific *chaînes opératoires* is preferable. If one works with fragments of specific *chaînes opératoires*, the overrepresentation of (e.g.) decorative and surface treatment techniques relative to roughing-out techniques skews the Wasserstein distance. Working with fragments of specific *chaînes opératoires* is possible but requires additional steps (see discussion on guided random path generation).

The third and last step in the comparison consists of interpreting the Wasserstein distance between body of knowledge A and B. What Wasserstein distance would warrant the conclusion that the potters involved did, or did not, learn the same production process? In order to tackle this question, we shall use a procedure which is analogous to *f*-statistics in genetics (cf. Patterson *et al.* 2012; Peter 2016) and polarisation in phylogenetics (Wiley and Lieberman 2011 pp. 10–1). A direct application of these algorithms is not possible because both procedures rely on the random nature of genetic mutations to convert the number of divergent mutations between two organisms or genomes into intervening generations (i.e. the more differences, the longer ago the last shared ancestor). The same principle does not apply to ceramic technology: mutations, or rather different technical choices, are not purely stochastic and recombination of 'different strands of knowledge' is common (see Section 3.2). However, the same base procedure can be applied. This procedure interpolates the relation between two groups through a comparison with a third, unrelated group.

The interpretation of Wasserstein distances requires at least three bodies of knowledge. For the first two bodies of knowledge, A and B, we do not know if potters shared knowledge (f.e. Funnel Beaker West and Corded Ware). For the third body of knowledge C, we know the potters learned a different production process than those who made the vessels in body of knowledge A (f.e. pottery production in modern India). All bodies of knowledge must consist of events within the sample space (hence the total *chaîne opératoire* must represent all specific *chaînes opératoires*). The Wasserstein distance between bodies of knowledge A and C then enables a conclusion about the Wasserstein distance between bodies of knowledge A and B (see Tab. 4.1; Fig. 4.2). After all, we know the distance between A and C is an indication of the learning and alternate choices potters would need to conduct a production process completely unrelated to what they themselves have learned.

In the above example, if the distance between Funnel Beaker West and Corded Ware *is equal to*, or *larger than*, the distance between Funnel Beaker West and pottery production in modern India, one may assume potters among Funnel Beaker West and Corded Ware communities never shared any knowledge. On the other hand, if the distance *is smaller than* that between Funnel Beaker West and pottery production in modern India, the results imply that the potters who made Corded Ware and Funnel Beaker West vessels did learn similar production processes (see Tab. 4.1). In this case, we can examine how much knowledge is shared between these potters. For example, by repeating step 3 with a fourth body of knowledge, such as specific *chaînes opératoires* from Funnel Beaker North, for

Outcome	Description	Interpretation
Wasserstein distance A-B < Wasserstein distance A-C	The variability in body of knowledge A requires fewer alternate choices and new techniques to produce the specific *chaînes opératoires* in body of knowledge B, than to produce the unrelated ones in body of knowledge C.	The potters who fashioned these vessels learned similar techniques.
Wasserstein distance A-B ≥ Wasserstein distance A-C	Within the variability of body of knowledge A, the amount of alternate choices and new techniques needed to produce the unrelated specific *chaînes opératoires* from body of knowledge C is equal to, or higher than those needed to produce body of knowledge B.	The potters who fashioned these vessels did not share technical knowledge.

Table 4.1: Potential outcomes of a probabilistic comparison of two bodies of knowledge. It is unknown whether the potters who made bodies of knowledge A and B shared knowledge, the potters who fashioned vessels in bodies of knowledge A and C are known to have learned different production processes. A comparison between A and C acts as a check for the relation between A and B.

which we know potters learned and used some of the same techniques as Funnel Beaker West (cf. Wiley and Lieberman 2011 p. 10 for a similar procedure in cladistics).

The sections above are a basic outline of a probabilistic approach to comparing specific *chaînes opératoires*. This outline suffices to interpret the analyses in this dissertation. However, there are also hints at a number of additional complexities in relation to the method. These complexities are discussed below in an in-depth account of the algorithms and assumptions involved in the comparisons.

Unlike Anything: Generating Random *Chaînes Opératoires*

As stated in the previous section, a comparison requires at least three bodies of knowledge. Two of these bodies of knowledge must have been fashioned by potters who did not learn similar techniques. To my knowledge, no such comparative data exists at present because the study protocol is novel. Therefore, this study relies on the random generation of specific *chaînes opératoires* to create bodies of knowledge for which we can be sure no shared knowledge exists with prehistoric potters. This section outlines the algorithm which generates these random specific *chaînes opératoires*. A more complex version of the same algorithm performs other tasks in this study (see Tab. 4.2). Hence the discussion of this version before the more complex version.

The generation of random *chaînes opératoires* is another affordance of the network representation of the total *chaîne opératoire*. This representation enables the use of a modified Markov chain (Duoc *et al.* 2010; Gagniuc 2017; see Tab. 4.3 for the algorithm). The modifications are necessary due to the complexities of ceramic production processes, but they also set the algorithm apart from a proper Markov chain.

Similar to a Markov chain, the algorithm for generating random specific *chaînes opératoires* constructs paths through a chance process. The algorithm takes the last node on the path and then determines the next node on the path by randomly selecting one of the neighbours of this node in the network. The algorithm appends this neighbour to the list and repeats the procedure (see Tab. 4.2). The random selection itself uses the *choice()* algorithm in Python (see Python documentation 2022 for the exact algorithm). The algorithm departs from a Markov chain in four ways: a fixed starting point, a memory,

Method	Random path generator	Guided random path generator
Network	Total *chaîne opératoire*	Total *chaîne opératoire*
Edge weight distribution	Uniform, all edge values equal 1.	Based on a body of knowledge, a seed, or uniform.
Edge weight modifier	None.	Depends on observed edges, and/or a set factor for non-observed edges.
Starting input	List with: 1) initial path state, 2) and initial node.	List with: 1) initial path state, 2) and initial node.
Step 1	List all neighbours of last node on list.	List all neighbours of last node on list.
Step 2	From the list in step 1, eliminate all neighbours which: 1) do not match path state in terms of tolerance, 2) already occur on the list.	From the list in step 1, eliminate all neighbours which: 1) do not match path state in terms of tolerance, 2) already occur on the list.
Step 3	Select one remaining neighbour at random, and append it to the list.	Select one remaining neighbour, and append it to the list. The odds of selection are proportional to the weight of the edge to the neighbour.
Step 4	Depending on new node, update path state. Return to step 1.	Depending on new node, update path state. Return to step 1.
Termination conditions	1) No neighbours available for selection in step 3: if path state is fired, save path as valid; if path state is not fired, discard path and re-start. 2) After passing a node for firing, equal chance of saving path as valid, or returning to step 1 when at step 4.	1) No neighbours available for selection in step 3: if path state is fired, save path as valid; if path state is not fired, discard path and re-start. 2) After passing a node for firing, chance of saving path as valid, or returning to step 1 when at step 4. Odds for either depend on observed post-firing techniques. Upon saving, split path in node pairs and adjust the weight of the respective edges by +1.
Output	Specific *chaînes opératoires*.	Specific *chaînes opératoires* and/or an edge weight distribution.

Table 4.2: Overview of base data, procedures, and output of two algorithms for the random generation of specific *chaînes opératoires*.

loop prevention, and a set of termination conditions. All four changes relate to ceramic technology, and in particular to syntax.

Firstly, the algorithm has a fixed starting point. Following the discussion in Section 4.1, all ceramic production processes start with the node 'wet clay', and in the hydric state 'wet'. Therefore, the algorithm starts with a list with two pieces of information. The first is a path state, namely 'wet' (see below), and the second is the starting node 'wet clay'. The selection of further nodes for the list is subject to the other three modifications.

The second modification relative to Markov chains relates to hydric states. Given the structuring role of the hydric states in ceramic production (see Section 4.1), the path has a separate variable which records its present hydric state. This variable is initially set to 'wet' (see above). The algorithm updates this variable when the nodes for 'drying to leather-hard', 'drying to dry', and 'firing' are appended to the list. The variable for the hydric state also plays a role in the selection of neighbours. Prior to randomly selecting a neighbour, the algorithm screens all candidates and removes all neighbours which do not feature the current hydric state of the path in their tolerance (see Tab. 4.2). This modification prevents violations of the hydric state in randomly generated paths: if the path state is 'wet', only techniques with the tolerance 'wet' can be selected as next nodes on the path.

The third modification governing the selection of nodes is loop prevention. Prior to selecting a neighbour, the algorithm removes all candidate nodes which are already in the path from the list of available neighbours (see Tab. 4.2). This prevents the algorithm from continuously selecting the same small set of nodes in a well-connected set.

Loop prevention is artificial, because loops do occur in ceramic technology (see examples below). However, most of these loops are a matter of splitting or lumping. For example, one can count the application of each coil as a distinct instance of coiling (see Sofaer 2018 on De La Fuente 2011), or group all of these actions without losing information.

There are also loops for which lumping is more problematic. For example, a potter may apply simple incisions to a vessel in a wet state, and later on return to add new incisions while the clay body is dry. In this example, loop prevention is definitely artificial, but it is preferable over allowing for loops for two reasons. Firstly, allowing for loops would only introduce more artificial decisions. It would necessitate rules which prevent the algorithm from ending up in an endless loop. In the above example, one would need to specify the maximum number of times simple incisions can occur in a specific *chaîne opératoire*, as well as the conditions in which this maximum number (or any other number) of loops is allowed. As such, one artificial decision to prevent loops is preferable over the many artificial decisions needed to allow for them.

In addition, allowing for loops would result in long, unwieldy randomly generated *chaînes opératoires*, because the chance of randomly selecting a node which advances the production process (such as drying or firing) among many, well-connected nodes for decorative techniques is small. These long randomly generated specific *chaînes opératoires* might end up being incomparable to human made ones, simply because potters have things to do other than endlessly applying decorations. As such, loop prevention is also preferable in terms of the computation and accuracy of the outcomes. However, loop prevention remains a choice between two evils.

The last modification of the algorithm relative to Markov chains is the implementation of termination conditions (see Tab. 4.2). The algorithm terminates a path under two conditions. The first condition is a lack of neighbours available for the selection procedure. In this case, the algorithm checks if the path state is 'fired' and stores the path if this is the case or discards the path if this is not the case. Secondly, after passing a node for firing, the algorithm applies a check prior to each new iteration (see Tab. 4.2). The check has an equal chance of terminating and storing the path or proceeding with a new iteration. Similar to loop prevention above, the aim of this change is to render the outcomes of random path generation more comparable to actual specific *chaînes opératoires*. Without the termination conditions, the algorithm would always produce paths with all post-firing treatments and only terminate after running out of nodes to append.

To return to the over-all algorithm, the fixed starting point, influence of hydric states, loop prevention, and termination conditions all act as a 'memory' during random path generation. Such a memory is a departure from Markov chains (cf. Duoc *et al.* 2010; Gagniuc 2017), but vital for generating credible specific *chaînes opératoires*.

All of the above modifications relate to a single principle, namely the distinction between sound, valid, and invalid paths. What do these terms mean? The total *chaîne opératoire* allows for a great number of paths. These paths fall into either of the above three categories on the basis of the following criteria. Sound and valid paths do not violate the syntax of ceramic production: these paths follow the hydric states in the order from high to low humidity, do not apply techniques suited for a particular hydric state outside of this state, and do not form endless loops. In other words, valid and sound paths are credible specific *chaînes opératoires*. They would work if a potter tried them. The difference between valid and sound paths lies in verification: valid paths follow the above rules, sound paths follow the rules and are attested within the ethnographic or archaeological record. Invalid paths on the other hand, are possible within the network but violate one or more of the above rules, and would therefore be unworkable as ceramic production processes. As such, the purpose of the modifications is to improve the efficiency of the algorithm. The modifications lead to the exclusion of invalid paths, and the generation of valid paths which are suitable for comparisons against sound paths.

It is possible to create a network in which only valid paths exist. However, this network is far less efficient and elegant than the network in Fig. 4.1. The network would essentially be a directed a-cyclical graph in which each node is a branching point for all possible valid paths which feature that node in that position. This graph essentially repeats all possible future branches for each branching point, resulting in a large, unwieldy network. The elegance and efficiency of the network in Fig. 4.1 lie precisely in the compression of all these alternate paths, but come at the cost of additional limitations on the algorithm (see above).

Why create an algorithm to generate random paths? Bodies of randomly generated specific *chaînes opératoires* are an ideal control group when checking for shared knowledge between archaeological datasets. A collection of randomly generated *chaînes opératoires* can, by definition, not be based on the same learned information as a body of knowledge. Therefore, the distance of an observed body of knowledge to a random body of knowledge is always a measure of the distance to a completely different ceramic production process (see Tab. 4.1). Randomly generated specific *chaînes opératoires* fulfil this role as the ultimate out-group in Chapter 10.

However, the use of randomly generated specific *chaînes opératoires* also necessitates considerations about robusticity. The randomly generated specific *chaînes opératoires* differ with each iteration of the algorithm. Random path generation can, by chance, result in a set of specific *chaînes opératoires* which resemble a body of knowledge. Therefore, a large number of randomly generated specific *chaînes opératoires* is needed to ensure a representative outcome. Following recommendations by Ripley (1987 p. 116), each body of randomly generated specific *chaînes opératoires* encompasses 1,000 paths; more than 5 times the size of the largest body of knowledge with observed specific *chaînes opératoires* (see Ch. 5). This volume ensures adequate reflection of randomness without excessive cost in terms of computation.

Random path generation enables the creation of bodies of knowledge which do not resemble extant data. In some instances, the opposite procedure may be necessary: the generation of paths which closely resemble a body of knowledge. For example, such a procedure can buffer sparse datasets. The next section is an outline of an algorithm for such a procedure.

Simulation and Guided Random Paths

The probabilistic approach has a second crucial affordance: the ability to simulate paths based on observed data. This application features prominently in Chapter 10. The algorithm involved is a more complex variant of random path generation. The complexity stems from the incorporation of edge weights in the selection of neighbours and the ability to perform a recursive process. This algorithm is referred to as guided random path generation or simulation.

The crucial modification occurs in step 3 of the algorithm (see Tab. 4.2). In guided random path generation, the chance of selecting a neighbour is proportional to the weight of the edge to this neighbour. Let us recall the edge weight stems from observed specific *chaînes opératoires* in a body of knowledge (see above). As a result, guided random path generation replicates the choices of potters to combine certain techniques. Furthermore, it is possible to adjust these edge weights or manipulate certain edge weights so as to increase or decrease the error rate of this replication process.

The second change to the base algorithm for random path generation occurs in the termination of the paths. The algorithm for guided random generation does not only store valid paths but can break up the valid path into node pairs and increase the weight of the corresponding edges by one (see Tab. 4.2). As such, the simulation can become recursive: the outcome of each iteration feeds into subsequent iterations.

Both changes enable guided random path generation to exhibit complex behaviour. This mode of path generation is able to perform various tasks. For present purposes, it serves to reconstruct bodies of knowledge from sparse data. As mentioned above, datasets with complete specific *chaînes opératoires* are rare. However, there are studies which describe fragments of specific *chaînes opératoires* for a given region and period. Guided random simulation can reconstruct complete specific *chaînes opératoires* from these fragments by joining the most prevalent (i.e. most likely) path segments together. This procedure is applied in Chapter 10 to simulate (i.a.) a dataset for Funnel Beaker North, which in turn helps to interpret the Wasserstein distance between Funnel Beaker West and Corded Ware (see above). Guided random path generation could also help to reconstruct complete specific *chaînes opératoires* from highly fragmented ceramics.

Lastly, guided random path generation opens up several new opportunities to apply computer sciences in studies of ceramic technology. For example, it could also simulate developments in ceramic technology by running multiple iterations with a given error rate. As such, this algorithm may prove crucial for studies of long-term developments and evolutionary trajectories in ceramic technology.

4.3 Final Considerations: Why Go Probabilistic?

This chapter is an outline of a probabilistic comparison of ceramic technology in two steps. The first step consists of envisioning ceramic production processes as paths (specific *chaînes opératoires*) through a network (the total *chaîne opératoire*). In the second step, these paths and the network are reconceptualised as analogues to a probability space. This analogy enables a comparison of the variability within bodies of knowledge possible, but also the generation of random and guided random specific *chaînes opératoires* which can serve as a check for shared knowledge. As such, this probabilistic method is tailored for the purposes of this study (see Ch. 3). However, the method also has advantages over current methods for comparing ceramic technology and can complement these approaches.

The most common method to compare specific *chaînes opératoires* is abductive in nature. The term abductive refers to the reconstruction of a single, narrative account of the *chaîne opératoire* from large numbers of pottery fragments and vessels. These narrative accounts list the techniques applied during each stage of the *chaîne opératoire*, and these lists can then be compared (see protocols in Gosselain 2018; Roux 2019a, e.g. 2019b).

The primary advantage of the abductive approach over the probabilistic approach outlined above is flexibility. The narrative account does not restrict the analysis to a specific resolution (e.g. techniques) nor require complete specific *chaînes opératoires* to conduct comparisons. For example, one could have a detailed account for roughing out of the lower body based on one group of sherds and a more general account of roughing-out techniques for the upper body which draws on a different set of sherds. By contrast, the probabilistic approach works best with complete specific *chaînes opératoires* at a consistent level of resolution. Consequently, the abductive approach is more flexible than the probabilistic approach outlined here.

However, the flexibility of the abductive approach comes with three disadvantages relative to the probabilistic approach. The first disadvantage relates to variability. The abstract narrative account produced during the abductive approach poorly accommodates assemblage variability, especially in comparisons between assemblages. Instead, the use of the abductive approach tends to lead to the detection of binary oppositions in specific stages. For example, researchers might state that traces of coiling are either present or absent during the roughing out of the lower body. Such binary oppositions do not tell us whether particular technical choices are common or rare, nor do they allow for the possibility that potters might learn or be familiar with multiple techniques. The probabilistic approach, on the other hand, is capable of incorporating variability in a more nuanced, quantitative fashion as multiple alternate paths with different intensities through the same network. Moreover, it makes no assumptions about the (in)ability of past potters to learn a given technique: it only increases the distance as potters need to learn more new techniques.

The second disadvantage of the abductive approach is unclarity with regard to what is compared. The abducted narratives may be based on ceramic assemblages with different properties. For example, one assemblage may feature a large number of wall sherds with traces of surface treatment but few bases with traces of roughing out, and vice versa for another assemblage. Yet the abductive approach would present the same narrative for both assemblages, despite these differences. How do these two narratives relate at all then? The advantage of the probabilistic comparisons lies in the explicit, quantifiable assumptions, base data, and output. Every datapoint is a specific *chaîne opératoire* which follows from traces on vessels, and has a measurable impact on the outcomes of the comparisons. As such, the probabilistic approach enjoys better verifiability than the abductive approach. This advantage is crucial for a deep time perspective on phenomena as complex as knowledge (cf. Mesoudi 2011).

The last disadvantage is the inability of the abductive approach to capture the syntax of the production process. The syntax is a crucial element in the definition of the *chaîne opératoire*: the ordering of the techniques is as important as the techniques used (see Section 4.1). The abductive approach loses sight of the syntax because it splits the *chaîne opératoire* into discrete steps and stages, which are compared separately. By contrast, the probabilistic approach can capture and compare these orders of techniques as well as the techniques themselves because it departs from network analysis which is based precisely on the connections between entities (see Section 4.1).

To conclude, both comparative methods have different strengths. The abductive approach provides flexibility, whereas the probabilistic approach can provide nuanced, quantifiable assessments of variability, which takes the syntax of the production process into account. Despite the methodological differences, both methods are ultimately compatible. Chapters 9 and 10 apply the abductive and probabilistic methods, respectively, to compare Funnel Beaker West and Corded Ware bodies of knowledge. The outcomes of both comparisons are similar and even complementary (see Ch. 12, 13).