



Rethinking Use-Wear Analysis and Experimentation as Applied to the Study of Past Hominin Tool Use

João Marreiros^{1,2,3}  · Ivan Calandra¹  · Walter Gneisinger¹ · Eduardo Paixão^{1,2,3} · Antonella Pedergrana¹ · Lisa Schunk^{1,2}

Published online: 4 May 2020
© The Author(s) 2020

Abstract

In prehistoric human populations, technologies played a fundamental role in the acquisition of different resources and are represented in the main daily living activities, such as with bone, wooden, and stone-tipped spears for hunting, and chipped-stone tools for butchering. Considering that paleoanthropologists and archeologists are focused on the study of different processes involved in the evolution of human behavior, investigating how hominins acted in the past through the study of evidence on archeological artifacts is crucial. Thus, investigating tool use is of major importance for a comprehensive understanding of all processes that characterize human choices of raw materials, techniques, and tool types. Many functional assumptions of tool use have been based on tool design and morphology according to archeologists' interpretations and ethnographic observations. Such assumptions are used as baselines when inferring human behavior and have driven an improvement in the methods and techniques employed in functional studies over the past few decades. Here, while arguing that use-wear analysis is a key discipline to assess past hominin tool use and to interpret the organization and variability of artifact types in the archeological record, we aim to review and discuss the current state-of-the-art methods, protocols, and their limitations. In doing so, our discussion focuses on three main topics: (1) the need for fundamental improvements by adopting established methods and techniques from similar research fields, (2) the need to implement and combine different levels of experimentation, and (3) the crucial need to establish standards and protocols in order to improve data quality, standardization, repeatability, and reproducibility. By adopting this perspective, we believe that studies will increase the reliability and applicability of use-wear methods on tool function. The need for a holistic approach that combines not only use-wear traces but also tool technology, design, curation, durability, and efficiency is also debated and revised. Such a revision is a crucial step if archeologists want to build major inferences on human decision-making behavior and biocultural evolution processes.

✉ João Marreiros
marreiros@rgzm.de

Keywords Functional analysis · Traceology · Use-wear analysis · Experimental archeology · Standards · Protocols · Methods

Introduction

From the earliest manifestations of technology in modern societies, technologies have enabled humans to adapt, express, and communicate (Foley and Lahr 2003; Holdaway and Douglass 2012; Morgan et al. 2015; Stout et al. 2011). At the same time, technologies likely reflect symbology, ethnic, and social status patterns, allowing the recognition and interpretation of different traits in the evolution of human behavior (d’Errico et al. 2003; Henshilwood et al. 2001). Technological changes were and still are a key factor in the dynamics of human societies, raising questions such as how and why those changes took place in earlier populations, and how people were able to survive and expand based on their technological traditions, innovations, and novelties (Brown et al. 2012; Klein 2000; Marean et al. 2007; McBrearty and Brooks 2000).

Since the Early Pleistocene—the longest and most defining period of our history—technology has been characterized by the use of tools made from different materials, such as stone, bone, and antler. Investigating how past hominins produced, designed, and used their tools—including possible variation through time and space—is therefore one of the key areas of research for understanding the evolution of human behavior (Ambrose 2001; Odell 2000). Over the last decades, archeological research has focused on diverse aspects of prehistoric tool-related behaviors of past humans, such as the procurement of raw materials (Andrefsky 1994, 2009; Brantingham 2003), diversification of manufacturing techniques (Kuhn 1992), the manipulation of physical proprieties (e.g., heat treatment) (Domanski et al. 1994; Domanski and Webb 1992; Mackay et al. 2018; Schmitt et al. 2013), and actual tool use (Keeley 1974). To determine tool use, use-wear analysis studies the characteristic patterns of macroscopic and microscopic traces left on tools’ surfaces after being used (Fullagar 2014). The formation processes of different types of use-wear traces have been shown to vary in response to changes in raw material, tool morphology, design, and tool function (action performed on a given worked material) (Astruc et al. 2003; Kamminga 1979; Lerner et al. 2007).

Thus, the current state-of-the-art in functional analysis heavily relies on experimental reference collections that aim to replicate past potential uses and understand the formation process of identifiable diagnostic traces, to which the archeological material can be compared (Coles 1979). Nevertheless, although inferring tool use was embedded in archeological research very early, the history of traceological studies is marked by several conceptual shifts and methodological improvements (Bicho et al. 2015; Dunmore et al. 2018; Marreiros et al. 2015). The discipline has faced many adjustments and developments according to changes in its methodological and theoretical scope and in the main research questions (Longo and Skakun 2008, Evans et al. 2014; Stemp et al. 2015; van Gijn 2014). As a result, much recent discussion has taken place about terminology and definitions of the discipline, most likely reflecting the continual improvement of different methodological protocols (Bradfield 2016). From our perspective, the different terminologies also reflect different objectives and scales of analysis (Marreiros et al. 2015). Thus, the term traceology refers to the study of all physical traces on an artifact’s surface, including use-wear, traces of production, non-utilitarian wear, and post-depositional

alterations. The study of wear traces resulting from post-depositional activities, for example, can be used to infer site formation processes (e.g., Burrioni 2002; Levi-Sala 1986; Asryan et al. 2014). Use-wear analysis refers to the study of traces present on the artifact's surfaces that result from the wear produced by human use. Both traceology and use-wear are ideally combined with residue analysis as a complementary source of evidence to infer the nature and character of the different traces (Pedergrana and Ollé 2017; Rots and Williamson 2004; Rots et al. 2015). On the other hand, and from our perspective, the term functional analysis assumes a wider approach—referring to a combination of methods and data—that allows researchers to assess tool design and use. In this case, use-wear and residue studies are part of the methods included in functional analysis, but not the only ones. Since the main goal of functional studies is to infer artifact function and design, features other than wear traces and residues are fundamental. Here, use-wear data should be combined with tool design, which includes tool morphology, intentional modification (e.g., retouch and resharpening), tool durability, and efficiency (Key and Lycett 2015; Key et al. 2015; Key 2020). When not used in combination with use-wear data, this approach is often called techno-functional analysis (Boëda 2001; Boëda and Auduze 2013; Lepot 1993). As the term indicates, this is based mainly on the morphological aspects of tools, with a special focus on the location and distribution of modified (i.e., retouched) and unmodified edges. These are in most cases supported by ethnographic observations or, occasionally, experiments. From our point of view, the so-called techno-functional method, based on tool design, can be used as an approach to promote some functional interpretations that, although being crucial to plan and design experiments, can only be tested when combined with use-wear and residue analysis (Bonilauri 2010; Pedergrana 2017). One of the best examples for the lack of such a complementary approach is the traditional descriptive lithic typology that often includes functional interpretations (e.g., knife, scraper, point). Unlike techno-functional analysis, traceological studies always include experimental replication and reference collections as a proxy to identify and interpret the archeological evidence (Adams and Adams 2009). Although it is beyond the scope of this paper to offer a new set of methods and/or theoretical orientations, here, we aim to discuss this methodological and conceptual trajectory. For this purpose, we address the major limitations of the discipline and propose some new research avenues, as well as highlight the importance of using standards and protocols in the discipline.

Traceology/Use-Wear Analysis: Brief Review of Methodological Developments

In the literature, it seems clear that early archeologists, intrigued by the nature of the artifact variability recovered from the archeological record, tried to evaluate tool use by past hominin populations (Binford 1973, 2019). Use-wear and traceology researchers would agree that since its beginnings, much has been said about the limitations of the discipline (Keeley 1974). When looking at the trajectory of the discipline, it seems clear to us that researchers invest a lot of effort on trying to improve their methods and tackling the continuous limitations of the discipline (Grace 1996). At the same time, such shifts can also be seen as the result of major changes in paradigms and theoretical agendas in archeological research (Piggott and Trigger 1991; Trigger 1989).

Initially embedded in the historical-culturalist approach that characterizes the archeological research of the late nineteenth and early twentieth centuries, the first attempts to study past tools aimed at understanding the technology (i.e., production) and use of tools as a reflection of past evolutionary steps in human behavior (Anderson et al. 1993). The variability of the archeological record was the main argument for the chronological and later regional organization of the different technological complexes in the Late Pleistocene and Early Holocene. Initially, functional interpretations were mainly based on the morphology and intentional modification (i.e., retouch) of tools. As such, retouched tools were interpreted per se as the crucial elements and used as *fossiles directeurs*, which led researchers to question these functional assumptions and interpretations (Bordes 1969; Bordes and de Sonneville-Bordes 1970).

These first attempts were followed by Semenov's systematic work (Semenov 1957), which is viewed by many as the pioneering work in the discipline of traceology (Phillips 1988). Widely known to Western European scholars at the end of the 1960s, after it was translated to English, his work was one of the first projects that established systematic experiments as a major proxy for identifying and interpreting the nature of the different wear traces. In fact, the relevance of experimental replication in Semenov's approach lies in the fact that it does not distinguish between technology and function. When both are combined, they can serve as a fundamental proxy for inferring not only the technological evolution but also the economic and social organization of past populations. Semenov also advocated the systematic use of microscopic observations. His predominant method was the so-called low-power approach, meaning the use of a low magnification stereomicroscope. It was also during this phase that the term traceology was coined, as a result of observations made during experiments, which highlighted the complexity of the mechanisms responsible for the different wear formation processes. These include both cultural (use-) and natural-wear processes.

In the early 1960's, the emergence of New Archeology, with its emphasis on natural and cultural phenomena as major factors for complex human social behavior (Binford 1962; Longacre 2010), led to a new emphasis on traceological studies. This phase was marked by testing the limitations pointed out by Semenov's work (Semenov 1957, 1964), which mainly concerned the difficulties of identifying and interpreting the diversity of wear traces when exclusively using a stereomicroscope (Keeley 1974; Odell 1975). To tackle this issue, several projects started introducing the so-called high-power approach, ultimately highlighting that both approaches are complementary and should be combined (Odell 2001). The high-power method refers to the use of high magnification microscopic equipment, mostly upright metallographic (light) microscopes. Mainly developed by Keeley (Keeley and Newcomer 1977), this approach showed that different types of wear traces could be recognized on tool surfaces (Hayden 1979). Experimental replication demonstrated that wear patterns were correlated with associated worked materials, allowing the categorization of diagnostic wear traces (Hardy and Garufi 1998; Hayden and Vaughan 2006). Researchers started realizing that such categorization was not always easy to interpret from archeological artifacts and on some raw materials. In other words, researchers realized that this qualitative analysis, while applicable to every raw material, requires that experiments be conducted for every different raw material type, constraining comparison between experimental and archeological assemblages (Burroni 2002; Evans and Donahue 2005).

While some researchers tried to improve the accuracy and reliability of the methodology through the development of numerous and more detailed experiments and blind tests,

others started arguing that use-wear analysis required new quantitative techniques (Grace et al. 1985; Keeley 1974; Schiffer 1979; Moss 1987; Bamforth 1988; Hurcombe 1988; Newcomer et al. 1986, Newcomer et al. 1988). Initially based on 2D images (Grace et al. 1985; Grace 1989), and later on 2D and 3D surface roughness measurements, this approach aims to create standardized criteria for variability within and between different types of wear traces. Such improvement was also made possible with the introduction of a wide range of imaging equipment and software, such as the tactile profilometer (Beyries et al. 1988), atomic force microscopy (Faulks et al. 2011; Kimball et al. 1995), interferometry (d’Errico and Backwell 2009; Dumont 1982), laser profilometry (Stemp and Stemp 2001, 2003), confocal microscopy (Evans and Donahue 2008; Evans 2014; Macdonald et al. 2018; Stemp et al. 2013; Stemp and Chung 2011), image analysis, and surface metrology software (Calandra et al. 2019a, b, c; d’Errico and Backwell 2009; Ibáñez et al. 2018; Martisius et al. 2018; Sahle et al. 2013). This variety of imaging equipment reflects the main aim of these approaches, which has focused on a quantitative identification and analysis of the different types of use-wear traces (e.g., polish and striations, see definitions below). Mainly applied to experimental replicas, from which the type of use-wear was already known, researchers applied different imaging techniques and methods at different scales. As a result, recent studies have shown that although quantitative methods can measure and identify diagnostic microscopic surface textures, much needs to be done in terms of data acquisition standardization so that data can be compared between experimental libraries. Analogous trajectories have also been observed in related disciplines, clearly illustrating continuous improvement in scientific fields (Calandra et al. 2019a).

This methodological discussion was also followed by a debate on the mechanical processes behind the formation of wear traces. Quantitative methods are not absolute but—like qualitative methods—require a suitable referential framework. This reference can be provided by comparison with ethnographic material and/or experiments. Thus, experimental design has also gained new attention and one could argue that it cannot be dissociated from the debate mentioned above. However, the same revision did not seem to have taken place with experimentation (Collins 2008). The properties of both the tool and worked material (e.g., hardness, density, grain size, elasticity, and surface roughness, to name a few) started to be seen as a major variable when trying to understand and identify the origins of the different types of wear. In fact, here, we argue that only a critical review and discussion of a combined approach consisting of (1) terminology, (2) microscopic methods, (3) material science (e.g., raw material properties and tribology), and (4) experiments can provide new insights and test the limitations of the discipline. Despite the fact that much has been done, one cannot deny that current state-of-the-art methods and protocols are still not mature and should, consequently, be improved. Here, we would like to address the actual limitations of the discipline by reviewing its four main issues and at the same time stress the need for implementing consensual standards and protocols in the field.

A Critical Review and New Insights

During the last decades, several researchers have pointed out limitations of the applied methods and protocols (Collins 2008; Dibble et al. 2017; Shea 1987; Dunmore et al. 2018). These limitations are often related to the following main issues: (1) archeological sampling and selection methods, (2) the lack of a clear and rigorous

experimental framework, (3) the lack of a standardized analytical framework, (4) subjective description of the wear traces, and (5) the absence of a holistic approach when interpreting tool use. As in any other discipline, methodological improvements do not necessarily reject previous work but rather aim to tackle fundamental limitations. While some methodological shifts are technology-based (i.e., microscopy techniques), others can be achieved by adopting clear standards and protocols already established in other fields. In this section, we review these topics and rethink how these can be tackled and improved.

Archeological Sampling and Selection Methods

In scientific research studies, sampling is a method for selecting a subset of individuals from within a statistical population (i.e. sample) to estimate characteristics of the whole population. Sampling is dependent on research questions and goals, and can therefore be performed by different methods. Sampling is fundamental in every field of scientific research. In archeology, sampling is always present, either on a field excavation or survey, artifact study, or data analysis (Nance 1994). For example, sampling strategies such as excavation units and sampling from different excavation areas reflect the main research questions of any given project, especially when the aim is to address questions such as a site's spatial organization. When compared with lithic technological and typological studies, traceology has been seen as a time-consuming method of analysis in which tools tend to be selectively sampled from the assemblage. Although this might not be a significant limitation when compared with other disciplines such as biology or chemistry (in which samples need long time preparation and analysis), researchers tend to use sampling strategies to subset the whole artifact assemblage. Observations made by analyzing this subset are then used to make inferences on the total assemblage. Traceological sampling strategies have been criticized as biased and therefore as one of the limitations of the discipline (e.g., Dibble et al. 2017). For example, in the analysis of lithic industries, studies have traditionally suggested the following technological organization: (1) debris, (2) debitage products, (3) cores, and (4) retouched tools. Here, because of their intentional modification and based on ethnographic studies, retouched tools have been interpreted as the most important tools for performing most of the tasks in the daily lives of past hominin populations (e.g., hunting spears). Based on functional interpretations, these different stone tools were often organized into different typologies used as *fossiles directeurs* for chronologies, geographies, and even techno-cultural contexts (Bordes 1988).

Seen as the only discipline for directly inferring tool use, use-wear analysis played a significant role in this discussion, allowing primary assumptions and functional interpretations to be addressed and tested (Semenov 1970). In contribution to this discussion, use-wear studies prioritized the analysis of retouched tools and often disregarded other technological classes, such as unmodified debitage products (i.e., unretouched flakes, blades, and bladelets) (Marreiros et al. 2015). However, recent research seemed to provoke a change in thinking concerning sampling, as traceological analysis of unmodified tools has shown the presence of use-wear traces, validating that these tools were also used. If analysis is limited to retouched tools, studies might be biased when interpreting the whole lithic assemblage and the nature of the human occupation of the site.

Sampling strategies must always reflect the research questions and main goals of any given study (Fig. 1). These in turn are highly dependent on the assumptions and/or

hypotheses to be tested. On the one hand, research questions and objectives are crucial when making decisions concerning sampling and applied methods; on the other, data should be collected independently of interpretations. When transferred to use-wear sampling methods, the strategy should focus on selecting tools not only based on preconceived functional categories or interpretations. In other words, one should not only select tools which are assumed to have been used, e.g., retouched tools as mentioned above. In lithic studies, for example, this can be easily done by sampling elements that represent all technological categories present in the lithic assemblage (Fig. 1, scenario a).

On the other hand, if, for instance, the research question is to analyze a specific morpho-type such as a group of lithic pointed blades, it makes sense that all elements that represent this category be selected, but the study should in addition also include other morpho-types as outgroups for comparison (Fig. 2, scenario b). As data from the study of the sampled tools aim to generate inferences on the entire assemblage, a study that only includes pointed blades cannot assess whether other tools show the same wear traces or not. In other words, from our perspective, when studies include only one type of tool and argue that this tool has been used in a specific motion, there is a risk that other excluded tools have been used to do the same. If this is the case, the interpretation of the site would be biased. These supplementary results may, however, be fundamental to further interpretations.

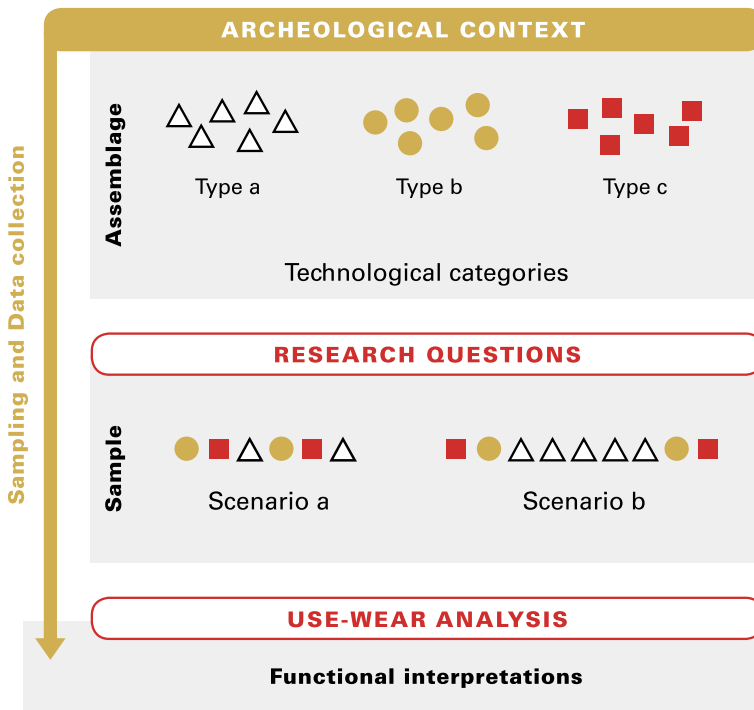


Fig. 1 Schematic representation of the different sampling and data collection strategies in functional studies

Traceology Experimental Framework

Experimental Archeology as a Scientific Method

Archeologists primarily aim for a comprehensive understanding of human behavioral evolution through inferring and reconstructing all past processes that shaped the evidence found in the archeological record. Archeologists' assumptions are mainly based on definitions, concepts, and interpretations arguably created from ethnographic or empirical observations and hypothesis testing (Binford 2019). Therefore, when analyzing the archeological record, evidence used to understand and reconstruct the evolution of human behavior heavily relies on experimental replication of potential past human activities (Miller and Hayden 2006; Eren et al. 2016; Lin et al. 2018). In this sense, it seems clear that if archeological research is to follow the scientific method, experimental archeology should be viewed as being in no way separate from other archeological sciences (Lin 2014; Lin et al. 2018). As a scientific "hypothetical-deductive" method, experimentation in archeological research has been used to test hypotheses and assumptions that are based on archeologists' observations and interpretations of the archeological record. This is clearly evident in technological studies, in which archeological experiments, included in the category of processes and functional experiments (Outram 2008), provide a crucial proxy for understanding and interpreting how different tools were produced, designed, and used in the past. As such, archeologists aim to run hypothesis-driven studies and acquire data independent from assumptions which are formulated through repeated empirical observations.

In experimental hypothesis testing, the conclusion is never proven to be true even if the tested premises are valid, and the predictions (i.e., tested assumptions, null hypothesis) are fulfilled. In fact, the final result can only be proven to be false and temporarily accepted as valid. This may be because scientific theories are based on observed effects (effect-causation relation), and here, many causes can lead to the same observed effects (equifinality; Eren et al. 2016). This phenomenon is amplified when experimental replications involve human behavior, which is known for being complex and idiosyncratically constrained.

In recent decades, several researchers have discussed the major limitations of experiments in archeological research, which can be summarized and categorized as follows: (a) lack of clear research questions, including hypothesis and assumptions to be tested, (b) alternative hypotheses are rarely tested, (c) insufficient details on the included material and methods, (d) lack of clear identification, organization, and definition of the control and manipulation of the different variables, (e) the coexistence of confounding variables, which refer to the presence of multiple variables included in the experiment at the same time, and (f) dominance of qualitative methods rather than quantitative (Collins 2008; Eren et al. 2016; Lin 2014). While these limitations in experimental replication should be tackled in order to improve the discipline, others can only be attributed to different hierarchical levels of experiments, these being dependent upon the main goal and research questions, which deeply influence the experimental organization, planning, and design. Organizing and planning experimental design are fundamental, not only to distinguish between strong and weak abductive inferences but also as a basis for others to use. From a scientific perspective, experiments should use structured and replicable processes to test the validity of a hypothesis. In this sense,

experiments are considered to be reliable (high internal validity) once they can be repeated, and their parameters and variables can be controlled and manipulated as needed and in multiple ways.

The importance of experimental organization and design has recently been discussed in depth in several highly important studies for inference validation and archeological interpretations (Ferguson and Neeley 2010). In this debate, researchers highlight the need to improve the way archeologists plan and design experiments, suggesting different levels of experimentation. These experiments, although complementary and crucial when constructing analogical inferences, should have different goals and seek different observations and types of data. In other words, the different levels of experiments are not distinguished by how the experiments are conducted but rather by which questions they aim to address.

Researchers have highlighted a distinction between two types of archeological experiments identified as: (1) “actualistic,” “pilot,” or “exploratory” and (2) “controlled,” “second generation,” or “laboratory” experiments. Due to their nature and aims, the most traditional approach, “actualistic,” “pilot,” or “exploratory” experimental studies (sensu Eren et al. 2016) mainly involve qualitative methods and a lower degree of variable control. These experiments are basically characterized by replicating human actions and, therefore, the use of mechanical apparatus is very limited. On the other hand, the so-called “controlled,” “second generation,” or “laboratory” experiments refer to experiments seeking to test individual variables and the interaction between them. In this case, these experiments are sometimes conducted by mechanical devices, through which many variables can be controlled and measured (Eren et al. 2016; Lin et al. 2018). As stressed by others, these two different categories do not exclude each other and do not reflect different research quality levels; rather, it is clear that both approaches are crucial and fundamental and should be complementary when designing a research project and addressing research questions. For example, the term “controlled experiments,” which implies that variables are controlled during the experiment, seems to be applicable to both types of experiments. Although some variables are always controlled or measured, “controlled conditions” have different meanings in both cases. As such, in our opinion, the terms first and second generation experiments appear to be more appropriate, reflecting at the same time the different levels but also the complementary character of both approaches that should be combined when planning and designing experiments (Fig. 1). A detailed discussion of this topic has been rigorously addressed by colleagues and is beyond the scope of the present study. Based on this discussion, which has focused on lithic technology, the aim here is to redirect this debate towards use-wear studies by stressing the importance of different levels of experimentation and suggesting the need for experimental standards and protocols that when reviewed and implemented can help traceologists to develop new research questions and acquire more crucial data. In use-wear studies, experiments are intended to test assumptions about identifiable parameters of tool use through the application of a hypothesis testing framework (Eren et al. 2016). In total accordance with this review, when transported to use-wear studies, we propose below a more refined terminology (Fig. 1).

Context	Category	OBJECTIVES	DESIGN	SAMPLE PREPARATION	VARIABLE CONTROL	OUTCOME
Experimental planning Archeological object // Research questions // Testing hypothesis	1st GENERATION	<ul style="list-style-type: none"> • Test initial observations and assumptions • Identify of major variables • First inferences on tool use (e.g. efficiency, durability, suitability) • Generate new hypothesis 	<ul style="list-style-type: none"> • Replication of potential past technologies 	<ul style="list-style-type: none"> • Sample variability and homogeneity (e.g. raw material, tool design, contact material) • Archeological replicas 	<ul style="list-style-type: none"> • Comparable samples • Comparable movements • Defined result • Record variable 	<ul style="list-style-type: none"> • Major variables to be tested • Major activities to be tested • Initial observations on effect-causation • Reference collection
	2nd GENERATION	<ul style="list-style-type: none"> • Test effect-causation of major variables • Test physical principles • Understanding mechanical processes • Develop units of analysis and measure 	<ul style="list-style-type: none"> • Mechanical devices • Automated apparatus 	<ul style="list-style-type: none"> • Raw material homogeneity • Archeological replicas • Standard samples • Standard contact material 	<ul style="list-style-type: none"> • Raw material properties • Standardization of major variables • Control and manipulation of variables • Standard movements 	<ul style="list-style-type: none"> • Sole predictors • Pattern recognition • Reference collection
	3rd GENERATION	<ul style="list-style-type: none"> • Test "human" variability • Test pattern recognition against archeological-replica assemblages • Develop units of observation for the archeological record 	<ul style="list-style-type: none"> • Replication of potential past technologies • Blind-test • Multi-sensor systems 	<ul style="list-style-type: none"> • Sample homogeneity (e.g. raw material, tool design, contact material) • Archeological replicas 	<ul style="list-style-type: none"> • Standard samples • Comparable samples • Measure and record all variables 	<ul style="list-style-type: none"> • Test against archeological record and reference collection • Test pattern identification and recognition • Reference collection

Fig. 2 Diagram representing the different categories included in the experimental planning. For each category, different aspects are discussed, concerning its objectives, design, sample preparation protocols, variable control, and outcome

First Generation

The first generation experiments are fundamental for generating new hypotheses and ideas and to help identify major variables that need to be tested individually as sole predictors of the final outcome. Also, these experiments are crucial for making preliminary observations on tool efficiency and durability, plus assessing the suitability of certain tools for performing a given activity. Since experiments consume both time and resources, these observations are also crucial for optimizing planning, costs, and timing, since major questions can be selected and prioritized, and experiments can be combined. In this case, we may safely assume that the first generation experiments provide important insights into our understanding of past technologies. Without such observations, our approach to technological studies would have followed different directions. The use of mechanical devices is often not included in the first generation experiments, where researchers replicate potential past technologies and aim to test replicas of archeological tools. Here, testing the variability of the archeological record is important, for example, by inferring the suitability of different raw materials for a given task. Nevertheless, although sample variability is tested, researchers should insure sample homogeneity, and therefore comparability. For example, if the goal is to infer tool use, since the contact material is not being tested, it should remain constant between experiments.

Second Generation

The second generation experiments involve more complex designs. This is due to the nature of different variables, with the aim of addressing more basic questions regarding some of the key properties of the archeological record, such as the diversity of rocks used to produce stone tools. In fact, this approach does not aim to provide a direct

explanation for the so-called big picture questions in a research project, but rather seeks the development of units of analysis and measure. Unlike the first generation experiments, the second generation experiments are built on principles that are not only independent of an archeologist's interpretations but are also built on uniformitarianism, i.e., physical principles which operate uniformly across space and time. They can be modeled independently of the research subject, aiming to detect patterns and build methodological and analytical units. These units of measure should always be independent from archeological interpretations, just as any data acquisition should be independent from data interpretation. These analytical units provide concrete data for the connections linking identified patterns and processes that can be used as proxies for inferring past human behavior. Here, quantification methods and techniques should be combined with qualitative descriptions in order to test and investigate previous assumptions and interpretations while at the same time improving accuracy, data quality, comparability, and enabling the evaluation of the final results. Thus, by following this approach, researchers should maximize their efforts to reduce subjectivity, which is inherent to qualitative analysis, and should preferentially include quantitative methods known for being more resistant to subjective interpretations.

As most experimental designs involve different variables and the first generation experiments seek the identification of major variables, these need to be tested individually in order to assess and evaluate their effect-causation and correlation with other variables in the experiment. Based on this evaluation, the second generation experiments often use mechanical or automated apparatuses which drastically reduce human variability, all the while controlling and manipulating system variables (Aldeias et al. 2016; Dibble and Rezek 2009; Eren et al. 2011; Iovita et al. 2013; Key 2016; Magnani et al. 2014; Schmidt et al. 2019; Astruc et al. 2003; Collins 2008; Lewis et al. 2011; Pflöging et al. 2019; Tomenchuk 1985; Calandra et al. 2020).

The same approach is often transferred to the selection of samples for the second generation experiments. It is very common to use standardized samples which imitate certain features of archeological artifacts. Thus—from our perspective—the second generation experiments are viewed by archeologists with a lot of criticism. This criticism is likely based on the fact that artifact production or the experimental replication of their usages is not related to human but rather to artificial activities. Nevertheless, the tested variables can be organized into two main groups: those related to (1) tool morphology and material properties and to (2) tool use per se. While variables related to tool use are more directly applicable to the archeological record, those related to the internal properties of the tool (e.g., edge angle and raw material) aim to test and understand the cause-effect relations linking these variables and mechanical processes.

In summary, from our perspective, compared with the first generation experiments, the second generation experiments do not usually aim to directly replicate human action or even real-life activities. Instead, they employ the so-called expert systems (mechanical devices) (Grace 1985). Expert systems create basic principles by isolating, controlling, and measuring the cause-effect of variables, from which a more detailed and precise understanding of the process can be deduced. Traceology is the study of wear traces resulting from mechanical actions (i.e., tribological phenomena that take place when two or more surfaces are in mechanical contact with each other); therefore, those uniformitarian mechanisms can be tested, and the causes for the observed effects can be

used to understand and interpret the processes behind the formation of different types of wear. These methods allow data analysis to measure effect sizes of the controlled variables on the dependent variable (e.g., roughness parameters), and therefore enable models to be built and tested using statistical analysis methods. Data analysis is also fundamental here. In order to optimize experiment time and sample preparation, the outcome data should be planned, and decisions regarding sample size and data analytical methods should be made from the beginning (Fisher 1935).

Third Generation

As these models are based on the uniformitarian processes and recognized patterns from the controlled experiments and standard samples, they need to be tested against archeological-like assemblages created and used by archeologists, incorporating human variability and varied actions but still independent of archeological interpretation. This step in the experimental replication can be defined as a third step, and accordingly, here, we call it a third generation experiment. Here, all generated models and patterns can be evaluated in a “naturalistic,” human made environment in which the human variability is included, and tools and actions similar to those identified in the archeological record can be tested. In order to test the identified patterns, these studies should involve both standardized and archeological-replica tools, where the latter can be tested against artifacts from the archeological record. The experimental design should not only involve the replication of potential past technologies, but human variability can also be tested and measured if technologies such as multi-sensor systems for gesture recognition are included (Key 2016; Pfleging et al. 2015; Williams-Hatala et al. 2018). Analysis of samples used in the third generation experiments can be conducted blind, meaning that the specific use of each tool will be unknown prior to data collection and the interpretation will be done via the combination of statistical modeling and traditional use-wear observations (Evans 2014; Rots et al. 2016).

Raw Material Characterization

Another important topic is the raw material variability present in the archeological assemblages. In the Late Pleistocene and Early Holocene research, differences in raw material properties are widely assumed to have been a major determining factor of variability in tool technology, production, and use (Delgado-Raack et al. 2009; Key et al. 2020; Nonaka et al. 2010; Odell 1981; Eren et al. 2014). Thus, understanding the nature of the major physical/mechanical properties is a pre-requisite for developing more reliable analyses to address prehistoric technology and tool function. This is especially true for use-wear studies where an understanding and a characterization of the physical variability of lithic raw materials are fundamentally linked to the formation and interpretation of use-wear traces (Lerner 2014). As an example, by identifying and measuring stone or bone properties likely to influence polish formation (such as surface hardness and roughness), greater insight can be gained into the nature of use-wear formation, thereby improving interpretations. In lithic studies, systematic quantification of fundamental properties of stone have led to advances in other areas of stone tool analysis such as raw material procurement (e.g., sourcing), manufacturing techniques (i.e., knapping), artifact morphology (i.e., typology), the manipulation of physical

properties (e.g., heat treatment), and actual tool-use (e.g., edge durability) (Andrefsky 2009; Mackay et al. 2018; Schmidt et al. 2016; Key et al. 2020; Schoville et al. 2017). These properties are also known to be responsible for the nature of different characteristic patterns of polish and damage and, therefore, need to be taken into consideration when experiments are designed (Evans 2014). Despite this, in functional studies, although material properties and their physical characteristics are often considered as a key factor, they are rarely tested and considered in the experimental design (Evans 2014).

A Complementary Experimental Workflow

As established earlier, the first step of the experimental planning is to define the research questions. These questions are open lines of research to which some assumptions have been made and from which hypotheses have been or can be made. Here, the hypotheses to be tested are mainly driven by previous observations, which are likely to result either from direct observation (inductive) or by testing the most probable explanation (abductive). Although a research question can be split into several hypotheses, these should be tested independently. The main distinction between the first and second generation experiments should be based on the research question to be addressed and the hypothesis to be tested.

For testing a given hypothesis, several types of variables need to be identified. In experimental design, variables can be organized into three main groups: independent, dependent, and confounding/random. Independent variables (inputs or causes) are those to be manipulated in order to measure their effects on the final result. Within the groups of independent variables, standardization is fundamental, and major parameters in the experiment should be constant unless testing them is an aim. Dependent variables (outputs or outcomes) are those “dependent” on the independent variable, which are expected to change based on the latter, and are therefore the ones being measured. The independent and dependent variables may be viewed in terms of cause and effect. If an independent variable is changed, then an effect is expected in the dependent variable. Confounding variables are all those variables that—due to experimental design constraints—are not identified as having an effect on the final result, or cannot be manipulated, controlled, tested, or measured. It should be highlighted that the variables can change their character according to the research questions, and testing hypothesis and can be manipulated accordingly.

This organization is fundamental in all generations of experiments and should be followed meticulously. As an example, experiments dealing with manual simulations and non-standard samples should try to standardize as much as possible the independent variable that is meant to be compared within and between experiments. Only such an approach can lead to inference validity and avoid equifinality, which is fundamental when comparisons are made and when pattern recognition is at the goal. Since all experiments include errors, all known possible sources of error should also be reported and taken into consideration when interpretations are made.

Traditionally used in use-wear studies, sequential experiments are intended to aid in the analysis of the cumulative development of wear formation. In this approach, tools are analyzed before and after cycles of use, with the aim not only of understanding how wear forms on a given raw material but also for deducing which variables in the

experiment (such as number of movements or cycles) are necessary for developing use-wear. Either using 2D or 3D methods, experimental results can be analyzed and compared if these experiments follow the same protocols and if the same location on the tool's surface is analyzed during the course of an experiment. Being able to relocate the area of interest on a sample is also essential for reproducibility in use-wear studies. Recently, a simple protocol to create a coordinate system directly on the experimental samples has been proposed (Calandra et al. 2019c) since visual relocation has limited applicability.

Finally—but not less importantly—experimental design should try to minimize the problem of equifinality (Ferguson and Neeley 2010) and circular reasoning. Given the fact that the same end results can potentially be reached by many processes, researchers should try to include alternative hypotheses that do not assess just the original assumptions and final end interpretations. In other words (as mentioned earlier in the discussion of sampling methods), if the research questions aim to understand if pointed blades were used as projectiles, then experiments should not just replicate projectile activities but also other actions that can be compared and rejected.

Lack of Standardization in Data Acquisition and Analysis

As traceology studies are highly dependent on microscopy, various pieces of imaging equipment have been introduced as a result of the development of new methods and improvements in the discipline and technological advances in microscopy. Budgetary constraints limit the range of microscopes available in every laboratory; this is even more critical when archeological collections cannot be transported to research laboratories. At the same time, the diversity of the archeological record—the wide variety of raw materials, different shapes and sizes of artifacts, plus the diversity of use-wear traces (Lerner et al. 2007; Lerner 2014)—makes it necessary to use different pieces of imaging equipment. In fact, the diversity of use-wear traces is not only due to diversity of raw material but also due to diversity of actions (movement, force, duration, worked material) and other non-use processes. All of these aspects make it clear why traceology, when compared with traditional lithic analysis, does not show a standardized analytical framework (Calandra et al. 2019a). To further improve methods, establishing a complete workflow and defining major standards and protocols on sample preparation, data acquisition, and data analysis seem a logical step.

Traceologists have been trying to establish a research workflow that integrates different scales of analysis and the identification of different types of use-wear traces (Borel et al. 2014; Adams 2014; Langejans 2011). This workflow is often combined with the technological and typological characterization of the artifact(s) studied, either lithics or bones. It usually includes sequential steps of analysis, aiming to infer different aspects: (1) preservation state, (2) macroscopic wear traces, (3) microscopic wear traces, and (4) residues. All these steps are categorized accordingly, and the use-wear position and distribution on the tool are recorded (see Fig. 3 as an example of a database structure applied for use-wear analysis).

This analysis is frequently supported by 2D digital photography of the use-wear traces, and an interpretation of the tool function can be proposed. In publications, researchers typically print pictures that best illustrate the wear traces, a selection of which sometimes under-represents the diversity of use-wear traces observed, but which

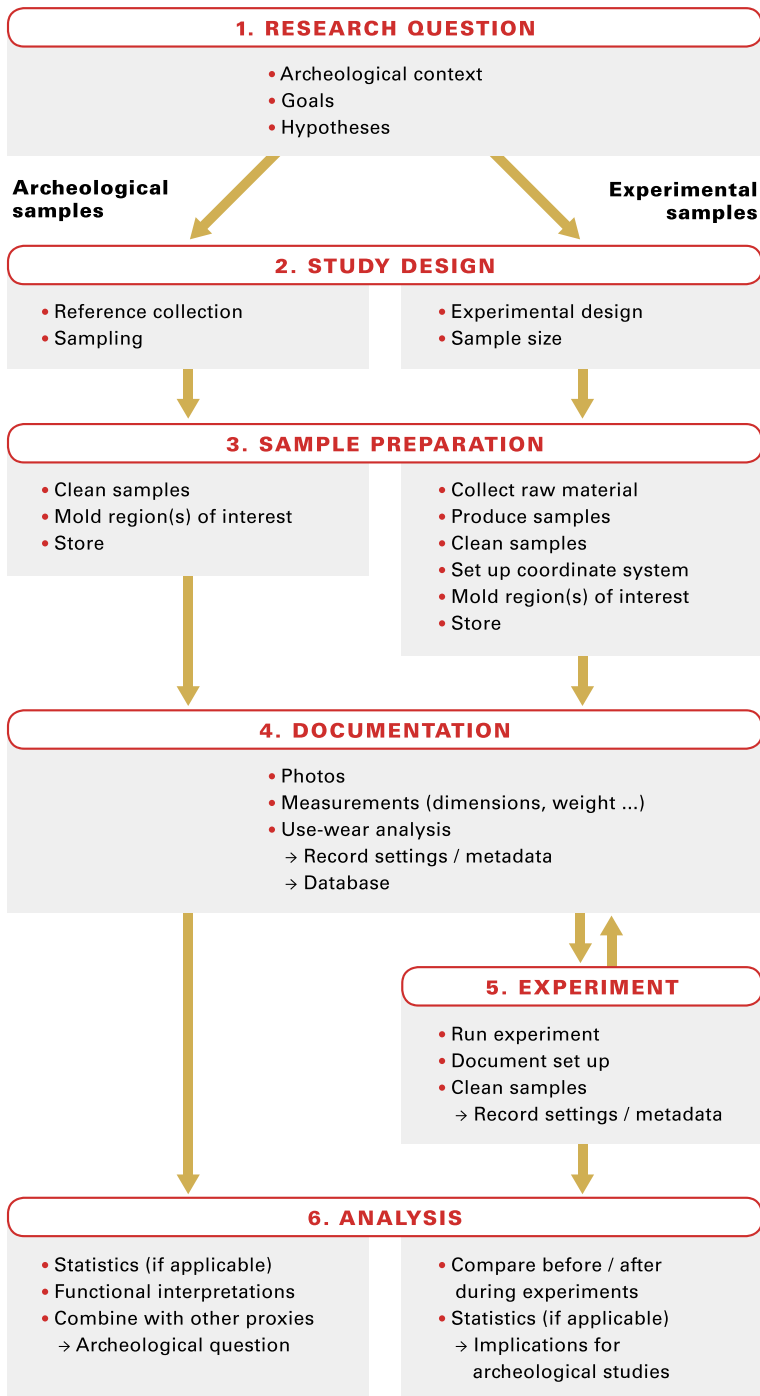


Fig. 3 Chart of the different methodological and analytical steps in the research workflow

does allow the wear traces to be easily and better understood by non-specialists. Although seemingly adequate, this practice often does not clearly illustrate the difference between the wear traces and their interpretations, especially when different worked materials are identified. Also, in the past, 2D images were sometimes of poor quality due to the lack of high quality cameras and extended depth of focus methods. Although these have been greatly improved in recent years, differences in resolution/scale between observations, often done through oculars, and published pictures with CCD/CMOS cameras, should be reported. Recent 3D microtexture methods focus on the quantification of the surface roughness (International Organization for Standardization 1996, 1997, 2012), allowing statistical analysis on the measured surface parameters. In addition, different wear traces can be compared between and within each other (Evans and MacDonald 2011; Galland et al. 2019; Ibáñez et al. 2018; Macdonald et al. 2018; Ollé and Vergès 2014; Stemp et al. 2019; Tumung et al. 2015).

Nevertheless, when using quantitative methods, reproducibility and repeatability are not automatic, and researchers need to be explicit, transparent, and extensive about their accounts of the methodology and settings. In order to secure repeatability and reproducibility in the discipline, standards and protocols are fundamental when studies aim for comparison and are even more important when quantification methods are applied (Calandra et al. 2019c). Quantification methods such as 3D microtexture analysis should also not be used alone. These methods complement the more traditional analyses and should be integrated in the analytical workflow (Evans 2014; van Gijn 2014). Qualitative low and high magnification observations are used to screen and identify the location of interest, in which use-wear is identified and documented. These areas are, in a second step, quantitatively analyzed for a more detailed characterization and comparison. Failing to do so, strict quantitative studies will be limited in their impact, as many researchers are not able to assess them and therefore integrate them in their projects (Calandra et al. 2019a). In parallel to this, additional standards and protocols need to be established, for example, concerning sample preparation procedures, including cleaning, archiving, and other processes such as molding and casting samples (Dunmore et al. 2018; Cnuts and Rots 2018; Langejans 2010; Macdonald et al. 2018). Although several protocols (e.g., cleaning solutions, Macdonalds and Evans 2014, Ollé and Vergès 2014) have been published lately that seem suitable for the visual identification of use-wear traces, it still needs to be evaluated how these procedures affect the surface microtopography.

Nature and Subjective Description of the Wear Traces

Traceology is a key discipline and currently the most reliable method for obtaining direct evidence on tool use by past hominins. The character of the discipline is based on a pattern recognition method when assessing the typology, location, and distribution of the different types of wear traces (Tringham et al. 1974). These studies became more important in archeological research when, based on experimental replication, researchers showed that wear patterns, especially polish, correlated with worked material (Keeley 1980; Haslam et al. 2009; Rots 2013). Traceologists have long tried to characterize observed wear patterns by describing not only their location and distribution but also their visual appearance. In the so-called qualitative approach, pattern recognition is based on the visual identification and interpretations of different features

(e.g., edge damage, diagnostic impact fractures, and striations) that characterize and categorize a given type of use-wear (McPherron et al. 2014; Schoville 2014). Among these different types, researchers have long argued that different types of polish correlate with different worked materials. Polish refers to a modified area on a tool surface, commonly described as a brighter or smoother area when observed with an optical microscope. These modifications are known to be the result of the interaction between different surfaces (i.e., tool and worked material) which causes a gradual removal or deformations of the natural surface.

Thus, the classification and quantification of polish became one of the fundamental questions to be addressed in use-wear analysis. During the last years the study of different types of polish has shown experimentally that such categories are likely related to many aspects such as the type of raw material, movement, number of cycles, and worked material. This means wear can be visually different, even on tools used to work the same contact material (Buc 2011). Taking into consideration that interpretations of tool use should not be based on polish only (discussed here later), the presence of use-wear polish already shows evidence for tool use, even if one cannot necessarily assess the worked material confidently.

Consequently, this approach has led to a broad variety of terms, mostly not understandable by others and lacking definition (e.g., wavy polish, here used as a basis for discussion) and/or carrying functional interpretations (e.g., cutting polish). This has led many researchers to highlight the two main limitations of the method: (1) a lack of understanding of the nature and influence of the mechanical process responsible for the formation of the different wear traces, and (2) unclear definitions and overlap of the so-called diagnostic wear traces, which often cannot be attributed to a specific worked material. In our opinion, although use-wear traces such as polish are not difficult to visually identify, their interpretation is rather more complex. This discussion most likely reflects the diversity of different types of wear traces (e.g., polish, striations, edge damage). This is not surprising since wear traces are the result of (1) human actions that are characterized by irregular movements even within the same action (e.g., cutting), meaning unsystematic movements, and (2) the contact of materials with different properties, even within the same type of rock (e.g., flint). That being said, much research needs to be done to fully understand the mechanisms behind the formation of polish, not least due to the different properties between and within tool raw materials (e.g., flint compared with quartzite and between the different types of flint). Perhaps, such wear trace types need to be tested and compared. On the other hand, such diversity has required site-based studies, which means that researchers often run experiments using the same raw material(s) that are represented in an archeological assemblage. In any typological approach, types are combined when assuming diagnostic characteristics or features. This is also claimed to be the case for the different types of wear polish, since researchers advocate the type of polish associated with bone work is significantly different from polish resulting from wood working.

Aiming to improve use-wear methods and to compare results from different experiments and from archeological sites, researchers have started proposing quantification methods that focus on the distribution and intensity of polish. These methods were based on digital image editing software (e.g., Fast Expert System, sensu Grace 1985) and GIS analysis (Dries 1998). This approach was again significantly improved when 3D microscopic quantification methods based on surface roughness measurements started to be

applied to the archeological studies (Dermdarsky and Ocklind 2001; Evans and Donahue 2008; González-Urquijo and Ibáñez-Estévez 2003; Kimball et al. 1995; Stemp and Stemp 2003). Nevertheless, such improvement still lacks an automated and standardized protocol of data acquisition and analysis, as well as experimental planning based on fundamental principles of reproducibility and repeatability, which would allow samples and data to be compared. In fact, the study of surface features has been undertaken and long established in a number of different fields in science and engineering (Calandra and Merceron 2016; Scott et al. 2005; Ungar and Evans 2016). In the recent past, these methods have been successfully harnessed in traceology to study artifacts by analyzing ISO parameters. This approach has now reached a high degree of sophistication in the field, for both experimental and archeological samples (Stemp 2014). Since quantification methods have comparison and frames of reference as their main goals, researchers have realized that it is fundamental to define standards and protocols that improve data repeatability, reproducibility, and comparability (Adams 2014; Calandra et al. 2019b; Dubreuil and Savage 2014; Evans et al. 2014). This is of particular importance since different pieces of equipment are used by various researchers, and different settings can be applied during the analysis (Stemp et al. 2013, 2018). Nevertheless, these methods will not change the fact that the nature of wear formation is diverse and is consequently represented in different types (Stemp et al. 2016). Similar discussions have also taken place, for example, in dental microwear studies, where researchers have shown that dental microwear patterns correlate with diet. In archeological samples, the variation in wear patterns emphasizes the need for protocols, standardization, and quantitative methods when documenting and measuring surface textures (Calandra et al. 2019b).

Absence of Holistic Approach when Interpreting Tool Function and Use

Reiterating, tool functional interpretations have long been based on tool morphology and intentional modification (i.e., retouch) and in analogy with the ethnographic record (McCall 2012). During the last decades, use-wear studies have shown that tool design does not always explain tool use (e.g., Evans 2014). Additionally, several studies have documented the unexpected presence of use-wear on certain tools (e.g., unmodified lithic implements such as cores, chunks, chips) or certain parts of a tool (e.g., broken edge), where by interpretation, a functional area has not been immediately recognized. This shows that humans use a huge panoply of elements in their daily life (Akoshima and Kanomata 2015). Researchers have consequently pointed to a lack of an integrative approach in use-wear studies when trying to reconstruct tool function and use (Rots et al. 2015; Rots and Plisson 2014). In this perspective, traceologists have only recently begun to incorporate their results into the technological and typological characterization of the tool and the whole assemblage (Zupancich et al. 2016; Sánchez-Yustos et al. 2017). From our understanding and implicit for our definition of functional analysis, data obtained by use-wear studies are a significant contributor to understanding tool use. At the same time, other aspects such as the production, durability, and efficiency of a given tool are crucial for addressing questions related to the understanding of fundamental mechanisms within the evolution of human behavior, such as cognitive, symbolical, and stylistic traits (Braun et al. 2008; Domínguez-Rodrigo and Pickering 2017; Pflöging et al. 2015).

Discussion and Conclusions

In what concerns studies of past human technology, it can be argued that interpretations of past human behavior depend exclusively on evidence from the archeological record, and archeological disciplines such as use-wear studies seek those direct pieces of evidence (Shea 2011). Based on experimental empirical observations and on analogy inferences (cause-effect), referential linkages fundamental for interpreting the archeological record can be deduced (Gaudzinski-Windheuser et al. 2018; Iovita et al. 2013). Furthermore, units of observation can bring new insights into interpretations of the archeological record and into understanding the different processes that shaped the evolution of human behavior.

Traceology is a fundamental field of investigation in archeological research. Our historical overview showed that methods and paradigms of the discipline have shifted over the past few decades, and the result is a well-established discipline that aims to combine different lines of evidence. In this paper, we provided a critical review and discussed major limitations of the discipline, at the same time highlighting and suggesting approaches for how these could potentially be tackled by specialists in this field. Our review shows that the nature of most of the limitations is no different from any other fields in archeological research, reflecting the trajectory and major methodological and theoretical debates in the history of archeological thought.

It is clear that the terminology of the discipline involves different scales of analysis and objectives, that each study addresses according to specific the research questions, and that methods and techniques should be adjusted accordingly. It is nevertheless important that researchers use a similar workflow and adhere to fundamental standards and protocols. Although quantification is an important step towards the standardization and calibration (sensu Evans et al. 2014) of the discipline, in the present paper, we highlight that such an approach should serve to categorize and evaluate our qualitative observations. Here, we point out that standards and protocols are important when—for example—sampling, cleaning tools, comparing tools during sequential experiments, and reporting all acquisition settings for different types of data. From our perspective, experimental organization and design are important methods within traceology. By discussing it and introducing a new category of experiments, we highlight how different generations of experiments aim at different goals and should be combined to achieve a more comprehensive understanding of all mechanisms.

Also, traceology or use-wear analysis is a significant part of functional analysis, and therefore, we argue that all evidence should be amalgamated in order to allow a more comprehensive interpretation of tool use. It is crucial that not only use-wear traces and residues are assessed (including their characteristics, location, and distribution on the tool surface) but also tool morphology, design, durability, and efficiency. From our perspective, a well-supported argument should combine all these different lines of evidence when experiments are planned and designed. Common standards and protocols should be implemented to allow for the evaluation of different variable causation, comparative results, and pattern recognition. By doing so, we believe that this approach will provide the insight from which significant assumptions about past human behavior can be made.

It should be said that the traditional descriptive terminology is still important, but when quantification and comparison need to be achieved, terms on unambiguous

definitions (i.e., standards) and protocols should be used. Therefore, it is fundamental to develop a corresponding “taxonomic” classification between qualitative and quantitative data. Since traceology relies on experimental reference collections and quantitative methods that are elementary for pattern recognition, studies should adopt not only fundamental standards and protocols but also methods of consistency and openness to their data and analysis. Recent studies have shown that in order to move traceology towards repeatability and standardization, acquisition and analysis settings need to be reported to improve reproducibility and open access in the discipline. Due to the fact that evidence originates from different sources and specific methodologies, we argue that standards and protocols should and need to be refined and improved. These should also be combined with a clear, detailed, and structured workflow for both experimental and archeological use-wear studies.

Acknowledgments The study presented here is the result of several insightful conversations with different colleagues, specialists, and non-specialists, on use-wear analysis and experiments. Among many, we would like to particularly thank Radu Iovita, Shannon McPherron, Telmo Pereira, Juan Gibaja, Nuno Bicho, Ariel Malinski-Buller, Olaf Jöris, and Harold Dibble.

Funding Information Open Access funding provided by Projekt DEAL. This research has been supported within the Römisch-Germanisches Zentralmuseum—Leibniz Research Institute for Archeology by German Federal and Rhineland Palatinate funding (Sondertatbestand “Spurenlabor”) and is publication no. 4 of the TraCEr laboratory.

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

References

- Adams, J. (2014). Ground stone use-wear analysis: a review of terminology and experimental methods. *Journal of Archeological Science*, 48(1), 129–138. <https://doi.org/10.1016/j.jas.2013.01.030>.
- Adams, W., & Adams, E. (2009). *Archeological typology and practical reality*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/cbo9780511558207>.
- Akoshima, K., & Kanomata, Y. (2015). *Technological organization and lithic microwear analysis: an alternative methodology* (Vol. 38). doi:<https://doi.org/10.1016/j.jaa.2014.09.003>.
- Aldeias, V., Dibble, H., Sandgathe, D., Goldberg, P., & McPherron, S. (2016). How heat alters underlying deposits and implications for archeological fire features: a controlled experiment. *Journal of Archeological Science*, 67, 64–79. <https://doi.org/10.1016/j.jas.2016.01.016>.
- Ambrose, S. (2001). Paleolithic technology and human evolution. *Science*, 291(5509), 1748–1753. <https://doi.org/10.1126/science.1059487>.
- Anderson, P., Beyries, S., & Otte, M. (1993). Traces et fonction: les gestes retrouvés: actes du colloque international de Liège, 8–10 décembre 1990.
- Andrefsky, W. (1994). Raw-material availability and the organization of technology. *American Antiquity*, 59(01), 21–34. <https://doi.org/10.2307/3085499>.
- Andrefsky, W. (2009). The analysis of stone tool procurement, production, and maintenance. *Journal of Archeological Research*, 17(1), 65–103. <https://doi.org/10.1007/s10814-008-9026-2>.

- Asryan, L., Ollé, A., & Moloney, N. (2014). Reality and confusion in the recognition of post-depositional alterations and use-wear: an experimental approach on basalt tools. *Journal of Lithic Studies*, 1, 9–32. <https://doi.org/10.2218/jls.v1i1.815>.
- Astruc, L., Vargiolu, R., & Zahouani, H. (2003). Wear assessments of prehistoric instruments. *Wear*, 255, 341–347. [https://doi.org/10.1016/S0043-1648\(03\)00173-X](https://doi.org/10.1016/S0043-1648(03)00173-X).
- Bamforth, D. (1988). Investigating microwear polishes with blind tests: the institute results in context. *Journal of Archaeological Science*, 15, 11–23.
- Beyries, S., Delamare, F., & Quantin, J.-C. (1988). Tracéologie et rugosimétrie tridimensionnelle. In S. Beyries (Ed.), *Industries Lithiques : Tracéologie et Technologie* (pp. 115–132). Oxford, UK.
- Bicho, N., Marreiros, J., & Gibaja, J. (2015). Use–wear and residue analysis in archeology. In J. M. Marreiros, J. F. Gibaja Bao, & N. Ferreira Bicho (Eds.), *Use-wear and residue analysis in archeology* (pp. 1–4). Cham: Springer International Publishing. doi:https://doi.org/10.1007/978-3-319-08257-8_1.
- Binford, L. (1962). Archeology as anthropology. *American Antiquity*, 28(2), 217–225. <https://doi.org/10.2307/278380>.
- Binford, L. (1973). Interassemblage variability - the Mousterian and the functional argument. In C. Renfrew & Research Seminar in Archeology and Related Subjects (Eds.), *The explanation of culture change: models in prehistory*. London: Duckworth. http://ls-tlss.ucl.ac.uk/course-materials/ARCL2037_42232.pdf
- Binford, L. (2019). *Constructing frames of reference: an analytical method for archeological theory building using hunter-gatherer and environmental data sets*. University of California Press.
- Boëda, É., (2001). Détermination des unités techno-fonctionnelles de pièces bifaciales provenant de la couche acheuléenne C3 base du site de Barbas I. In: Cliquet, D. (Ed.), *Les industries à outils bifaciaux du Paléolithique moyen d'Europe occidentale*. Actes de la Table Ronde internationale, Caen, 14-15 octobre 1999, Liège, ERAUL 98, pp. 51–75.
- Boëda, É., & Audouze, F. (2013). Techno-logique & technologie: une paléo-histoire des objets lithiques tranchants. @ rchéo-éditions.
- Lepot, (1993). Approche techno-fonctionnelle de l'outillage lithique moustérien : essai de classification des parties actives en termes d'efficacité technique. Application à la couche M2e sagittale du Grand Abri de la Ferrassie (fouille Henri Delporte). Master thesis submitted to the University of Paris-X-Nanterre (Université de Paris X-Nanterre, Département d'Ethnologie, de Sociologie comparative et de Préhistoire, Paris).
- Bonilauri, S., (2010). Les outils du Paléolithique moyen : une mémoire technique oubliée? Approche techno-fonctionnelle appliquée à un assemblage lithique de conception Levallois provenant du site d'Umm el Tiel ((Syrie centrale). PhD thesis submitted to the University of Paris Ouest-Nanterre La Défense (Université Paris Ouest Nanterre La Défense).
- Bordes, F., (1988). Typologie du Paléolithique ancien et moyen. CNRS (Originally published in 1950).
- Bordes, F. (1969). Reflections on typology and technology in the palaeolithic. *Arctic Anthropology*, 6, 1–29.
- Bordes, F., & de Sonneville– Bordes, D. (1970). The significance of variability in Paleolithic assemblages. *World Archeology*. <https://doi.org/10.1080/00438243.1970.9979464>.
- Borel, A., Ollé, A., Vergès, J. M., & Sala, R. (2014). Scanning electron and optical light microscopy: two complementary approaches for the understanding and interpretation of usewear and residues on stone tools. *Journal of Archaeological Science, Lithic Microwear Method: Standardisation, Calibration and Innovation*, 48, 46–59. <https://doi.org/10.1016/j.jas.2013.06.031>.
- Bradfield, J. (2016). Use-trace epistemology and the logic of inference. *Lithic Technology*, 41(4), 293–303. <https://doi.org/10.1080/01977261.2016.1254360>.
- Brantingham, P. (2003). A neutral model of stone raw material procurement. *American Antiquity*, 68(3), 487–509. <https://doi.org/10.2307/3557105>.
- Braun, D., Pobiner, B., & Thompson, J. (2008). An experimental investigation of cut mark production and stone tool attrition. *Journal of Archeological Science*, 35(5), 1216–1223. <https://doi.org/10.1016/j.jas.2007.08.015>.
- Brown, K., Marean, C., Jacobs, Z., Schoville, B., Oestmo, S., Fisher, E., et al. (2012). An early and enduring advanced technology originating 71,000 years ago in South Africa. *Nature*, 491(7425), 590–593. <https://doi.org/10.1038/nature11660>.
- Buc, N. (2011). Experimental series and use-wear in bone tools. *Journal of Archeological Science*, 38(3), 546–557. <https://doi.org/10.1016/j.jas.2010.10.009>.
- Burroni, D. (2002). The surface alteration features of flint artefacts as a record of environmental processes. *Journal of Archeological Science*, 29(11), 1277–1287. <https://doi.org/10.1006/jasc.2001.0771>.
- Calandra, I., & Merceron, G. (2016). Dental microwear texture analysis in mammalian ecology. *Mammal Review*, 46(3), 215–228. <https://doi.org/10.1111/mam.12063>.

- Calandra, I., Pedergrana, A., Gneisinger, W., & Marreiros, J. (2019a). Why should traceology learn from dental microwear, and vice-versa? *Journal of Archeological Science*, *110*, 105012. <https://doi.org/10.1016/j.jas.2019.105012>.
- Calandra, I., Schunk, L., Bob, K., Gneisinger, W., Pedergrana, A., Paixao, E., Hildebrandt, A., & Marreiros, J. (2019b). The effect of numerical aperture on quantitative use-wear studies and its implication on reproducibility. *Scientific Reports*, *9*(1), 6313. <https://doi.org/10.1038/s41598-019-42713-w>.
- Calandra, I., Schunk, L., Rodriguez, A., Gneisinger, W., Pedergrana, A., Paixao, E., Pereira, T., Iovita, R., & Marreiros, J. (2019c). Back to the edge: relative coordinate system for use-wear analysis. *Archeological and Anthropological Sciences*. <https://doi.org/10.1007/s12520-019-00801-y>.
- Calandra, I., Gneisinger, W., & Marreiros, J. (2020). A versatile mechanized setup for controlled experiments in archaeology. *STAR: Science & Technology of Archaeological Research*. <https://doi.org/10.1080/20548923.2020.1757899>.
- Cnats, D., & Rots, V. (2018). Extracting residues from stone tools for optical analysis: towards an experiment-based protocol. *Archeological and Anthropological Sciences*, *10*(7), 1717–1736. <https://doi.org/10.1007/s12520-017-0484-7>.
- Coles, J. (1979). *Experimental archeology*. London, UK: Academic Press.
- Collins, S. (2008). Experimental investigations into edge performance and its implications for stone artefact reduction modelling. *Journal of Archeological Science*, *35*, 2164–2170. <https://doi.org/10.1016/j.jas.2008.01.017>.
- d'Errico, F., & Backwell, L. (2009). Assessing the function of early hominin bone tools. *Journal of Archeological Science*, *36*(8), 1764–1773. <https://doi.org/10.1016/j.jas.2009.04.005>.
- d'Errico, F., Henshilwood, C., Lawson, G., Vanhaeren, M., Tillier, A., Soressi, M., et al. (2003). Archeological evidence for the emergence of language, symbolism, and music—an alternative multidisciplinary perspective. *World*, *17*(1), 1–70.
- Delgado-Raack, S., Gómez-Gras, D., & Risch, R. (2009). The mechanical properties of macrolithic artifacts: a methodological background for functional analysis. *Journal of Archeological Science*, *36*(9), 1823–1831. <https://doi.org/10.1016/j.jas.2009.03.033>.
- Demdarsky, M., & Ocklind, G. (2001). Some preliminary observations on subsurface damage on experimental and archeological quartz tools using CLSM and dye. *Journal of Archeological Science*. <https://doi.org/10.1006/jasc.2000.0646>.
- Dibble, H., Holdaway, S., Lin, S., Braun, D., Douglass, M., Iovita, R., et al. (2017). Major fallacies surrounding stone artifacts and assemblages. *Journal of Archeological Method and Theory*, *24*(3), 813–851. <https://doi.org/10.1007/s10816-016-9297-8>.
- Dibble, H., & Rezek, Z. (2009). Introducing a new experimental design for controlled studies of flake formation: results for exterior platform angle, platform depth, angle of blow, velocity, and force. *Journal of Archeological Science*, *36*(9), 1945–1954. <https://doi.org/10.1016/j.jas.2009.05.004>.
- Domanski, M., & Webb, J. (1992). Effect of heat treatment on siliceous rocks used in prehistoric lithic technology. *Journal of Archeological Science*, *19*(6), 601–614. [https://doi.org/10.1016/0305-4403\(92\)90031-W](https://doi.org/10.1016/0305-4403(92)90031-W).
- Domanski, M., Webb, J., & Boland, J. (1994). Mechanical properties of stone artefact materials and the effect of heat treatment. *Archaeometry*, *2*(August 1993), 177–208. <https://doi.org/10.1111/j.1475-4754.1994.tb00963.x>.
- Dominguez-Rodrigo, M., & Pickering, T. (2017). The meat of the matter: an evolutionary perspective on human carnivory. *Azania: Archeological Research in Africa*, *52*(1), 4–32. <https://doi.org/10.1080/0067270X.2016.1252066>.
- Dries, M. van den, (1998). Archeology and the application of artificial intelligence : case-studies on use-wear analysis of prehistoric flint tools. Doctoral Thesis. University of Leiden. <https://openaccess.leidenuniv.nl/handle/1887/13148>. Accessed 3 July 2019.
- Dubreuil, L., & Savage, D. (2014). Ground stones: a synthesis of the use-wear approach. *Journal of Archeological Science*, *48*(1), 139–153. <https://doi.org/10.1016/j.jas.2013.06.023>.
- Dumont, J. (1982). The quantification of microwear traces: a new use for interferometry. *World Archeology*, *14*(2), 206–217. <https://doi.org/10.1080/00438243.1982.9979861>.
- Dunmore, C., Pateman, B., & Key, A. (2018). A citation network analysis of lithic microwear research. *Journal of Archeological Science*, *91*, 33–42. <https://doi.org/10.1016/j.jas.2018.01.006>.
- Eren, M., Boehm, A., Morgan, B., Anderson, R., & Andrews, B. (2011). Flaked stone taphonomy: a controlled experimental study of the effects of sediment consolidation on flake edge morphology. *Journal of Taphonomy*, *9*(3), 201–217.

- Eren, M., Lycett, S., Patten, R., Buchanan, B., Pargeter, J., & O'Brien, M. (2016). Test, model, and method validation: the role of experimental stone artifact replication in hypothesis-driven archeology. *Ethnoarcheology*, 8(2), 103–136. <https://doi.org/10.1080/19442890.2016.1213972>.
- Eren, M., Roos, C., Story, B., von Cramon-Taubadel, N., & Lycett, S. (2014). The role of raw material differences in stone tool shape variation: an experimental assessment. *Journal of Archeological Science*, 49, 472–487. <https://doi.org/10.1016/j.jas.2014.05.034>.
- Evans, A., & MacDonald, D. (2011). Using metrology in early prehistoric stone tool research: further work and a brief instrument comparison. *Scanning*, 33(5), 294–303. <https://doi.org/10.1002/sca.20272>.
- Evans, A., & Donahue, R. (2005). The elemental chemistry of lithic microwear: an experiment. *Journal of Archeological Science*, 32(12), 1733–1740. <https://doi.org/10.1016/j.jas.2005.06.010>.
- Evans, A., & Donahue, R. (2008). Laser scanning confocal microscopy: a potential technique for the study of lithic microwear. *Journal of Archeological Science*, 35(8), 2223–2230. <https://doi.org/10.1016/j.jas.2008.02.006>.
- Evans, A. (2014). On the importance of blind testing in archeological science: the example from lithic functional studies. *Journal of Archeological Science*, 48, 5–14. <https://doi.org/10.1016/j.jas.2013.10.026>.
- Evans, A., Lerner, H., Macdonald, D., Stemp, W., & Anderson, P. (2014). Standardization, calibration and innovation: a special issue on lithic microwear method. *Journal of Archeological Science*, 48, 1–4. <https://doi.org/10.1016/j.jas.2014.03.002>.
- Faulks, N., Kimball, L., Hidjrati, N., & Coffey, T. (2011). Atomic force microscopy of microwear traces on Mousterian tools from Myshtylagty Lagat (Weasel Cave), Russia. *Scanning*, 33(5), 304–315. <https://doi.org/10.1002/sca.20273>.
- Ferguson, J., & Neeley, M. (2010). Designing experimental research in archeology: examining technology through production and use. *Ethnoarcheology*, 262, 29–95. <https://doi.org/10.4207/PA.2010.REV93>.
- Fisher, R. (1935). The logic of inductive inference. *Journal of the Royal Statistical Society*, 98(1), 39–82. <https://doi.org/10.2307/2342435>.
- Foley, R., & Lahr, M. (2003). On stony ground: Lithic technology, human evolution, and the emergence of culture. *Evolutionary Anthropology: Issues, News, and Reviews*, 12(3), 109–122. <https://doi.org/10.1002/evan.10108>.
- Fullagar, R. (2014). Residues and usewear. In J. Balme & A. Paterson (Eds.), *Archeology in practice: A student guide to archeological analyses* (2nd ed., pp. 232–263). Malden: Blackwell Publishing <https://ro.uow.edu.au/smhpapers/2745>.
- Galland, A., Queffelec, A., Caux, S., & Bordes, J. G. (2019). Quantifying lithic surface alterations using confocal microscopy and its relevance for exploring the Neanderthal-Châtelperronian association at La Roche-à-Pierrot (Saint-Césaire, France). *Journal of Archeological Science*, 104(January), 45–55. <https://doi.org/10.1016/j.jas.2019.01.009>.
- Gaudzinski-Windheuser, S., Noack, E., Pop, E., Herbst, C., Pflöging, J., Buchli, J., Buchli, J., Jacobs, A., Anzmann, F., Kindler, L., Iovita, R., Street, M., & Roebroeks, W. (2018). Evidence for close-range hunting by last interglacial Neanderthals. *Nature Ecology & Evolution*, 2(7), 1087. <https://doi.org/10.1038/s41559-018-0596-1>.
- González-Urquijo, J., & Ibáñez-Estévez, J. (2003). The quantification of use-wear polish using image analysis. First results. *Journal of Archeological Science*, 30(4), 481–489. doi:<https://doi.org/10.1006/jasc.2002.0855>.
- Grace, R. (1996). Review article on use-wear analysis: the state of the art. *Archaeometry*, 38(2), 209–229. <https://doi.org/10.1111/j.1475-4754.1996.tb00771.x>.
- Grace, R., Graham, I., & Newcomer, M. (1985). The quantification of microwear polishes. *World Archeology*, 17(1), 112–120. <https://doi.org/10.1080/00438243.1985.9979954>.
- Grace, R. (1989). Interpreting the function of stone tools : the quantification and computerisation of microwear analysis. **British Archaeological Reports International**.
- Hardy, B., & Garufi, G. (1998). Identification of woodworking on stone tools through residue and use-wear analyses: experimental results. *Journal of Archeological Science*, 25(2), 177–184. <https://doi.org/10.1006/jasc.1997.0234>.
- Haslam, M., Robertson, G., Crowther, A., Nugent, S., & Kirkwood, L. (Eds.). (2009). *Archeological science under a microscope* (Terra Australis Vol. 30). ANU Press.
- Hayden, B., & Vaughan, P. (2006). Use-wear analysis of flaked stone tools. *Man*, 21(3), 553. <https://doi.org/10.2307/2803118>.
- Hayden, B. (1979). *Lithic use-wear analysis*. Academic Press.
- Henshilwood, C. S., D'Errico, F., Marean, C. W., Milo, R. G., & Yates, R. (2001). An early bone tool industry from the middle stone age at Blombos Cave, South Africa: implications for the origins of modern human

- behaviour, symbolism and language. *Journal of Human Evolution*, 41(6), 631–678. <https://doi.org/10.1006/jhev.2001.0515>.
- Holdaway, S., & Douglass, M. (2012). *A twenty-first century archeology of stone artifacts* (Vol. 19). <https://doi.org/10.1007/s10816-011-9103-6>.
- Hurcombe, L. (1988). Some criticisms and suggestions in response to Newcomer et al. (1986). *Journal of Archaeological Science*, 15, 1–10.
- Ibáñez, J., Lazuen, T., & González-Urquijo, J. (2018). Identifying experimental tool use through confocal microscopy. *Journal of Archeological Method and Theory*. <https://doi.org/10.1007/s10816-018-9408-9>.
- International Organization for Standardization. (1996). ISO 4288 – Geometrical product specifications (GPS) – surface texture: profile method – rules and procedures for the assessment of surface texture.
- International Organization for Standardization. (1997). ISO 4287 – Geometrical product specifications (GPS) – surface texture: profile method – terms, definitions and surface texture parameters.
- International Organization for Standardization. (2012). ISO 25178-2 – Geometrical product specifications (GPS) – surface texture: areal – part 2: terms, definitions and surface texture parameters.
- Iovita, R., Schöneke, H., Gaudzinski-Windheuser, S., & Jäger, F. (2013). Projectile impact fractures and launching mechanisms: results of a controlled ballistic experiment using replica Levallois points. *Journal of Archeological Science*, 48, 73–83. <https://doi.org/10.1016/j.jas.2013.01.031>.
- Kamminga, J. (1979). The nature of use-polish and abrasive smoothing on stone tools. In *Lithic Use-Wear Analysis* (pp. 143–158).
- Keeley, L. (1974). Technique and methodology in microwear studies: a critical review. *World Archeology*, 5(3), 323–336. <https://doi.org/10.1080/00438243.1974.9979577>.
- Keeley, L. H. (1980). *Experimental determination of stone tool uses: A microwear analysis*. Chicago: University of Chicago Press.
- Keeley, L., & Newcomer, M. (1977). Microwear analysis of experimental flint tools: a test case. *Journal of Archeological Science*, 4(1), 29–62. [https://doi.org/10.1016/0305-4403\(77\)90111-X](https://doi.org/10.1016/0305-4403(77)90111-X).
- Key, A., & Lycett, S. (2015). Edge angle as a variably influential factor in flake cutting efficiency: an experimental investigation of its relationship with tool size and loading. *Archaeometry*, 57(5), 911–927. <https://doi.org/10.1111/arc.12140>.
- Key, A., Stemp, W. J., Morozov, M., Proffitt, T., & de la Torre, I. (2015). Is loading a significantly influential factor in the development of lithic microwear? An experimental test using LSCM on basalt from Olduvai Gorge. *Journal of Archeological Method and Theory*, 22(4), 1193–1214. <https://doi.org/10.1007/s10816-014-9224-9>.
- Key, A., Proffitt, T., & de la Torre, I. (2020). Raw material optimization and stone tool engineering in the early stone age of Olduvai Gorge (Tanzania). *Journal of the Royal Society Interface*, 17, 20190377. <https://doi.org/10.1098/rsif.2019.0377>.
- Key, A. (2016). Integrating mechanical and ergonomic research within functional and morphological analyses of lithic cutting technology: Key principles and future experimental directions. *Ethnoarcheology*, 8(1), 69–89. <https://doi.org/10.1080/19442890.2016.1150626>.
- Kimball, L., Kimball, J., & Allen, P. (1995). Microwear polishes as viewed through the atomic force microscope. *Lithic Technology*, 20(1), 6–28.
- Klein, R. (2000). Archeology and the evolution of human behavior. *Evolutionary Anthropology: Issues, News, and Reviews*, 9(1), 17–36. [https://doi.org/10.1002/\(SICI\)1520-6505\(2000\)9:1<17::AID-EVAN3>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1520-6505(2000)9:1<17::AID-EVAN3>3.0.CO;2-A).
- Kuhn, S. (1992). On planning and curated technologies in the Middle Paleolithic. *Journal of Anthropological Research*, 48(3), 185–214. <https://doi.org/10.1086/jar.48.3.3630634>.
- Langejans, G. (2010). Remains of the day-preservation of organic micro-residues on stone tools. *Journal of Archeological Science*, 37(5), 971–985. <https://doi.org/10.1016/j.jas.2009.11.030>.
- Langejans, G. (2011). Discerning use-related micro-residues on tools: testing the multi-stranded approach for archeological studies. *Journal of Archeological Science*, 38(5), 985–1000. <https://doi.org/10.1016/j.jas.2010.11.013>.
- Lerner, H., Du, X., Costopoulos, A., & Ostojca-Starzewski, M. (2007). Lithic raw material physical properties and use-wear accrual. *Journal of Archeological Science*, 34(5), 711–722. <https://doi.org/10.1016/j.jas.2006.07.009>.
- Lerner, H. J. (2014). Intra-raw material variability and use-wear accrual: a continuing exploration. *Journal of Lithic Studies*, 1(1), 165–186. <https://doi.org/10.2218/jls.v1i1.755>.
- Levi Sala, I. (1986). Use wear and post-depositional surface modification: a word of caution. *Journal of Archeological Science*, 13, 229–244. [https://doi.org/10.1016/0305-4403\(86\)90061-0](https://doi.org/10.1016/0305-4403(86)90061-0).
- Lewis, R., Tsoraki, C., Broughton, J., et al. (2011). Abrasive and impact wear of stone used to manufacture axes in Neolithic Greece. *Wear*, 271, 2549–2560. <https://doi.org/10.1016/j.wear.2010.12.074>.

- Lin, S., Rezek, Z., & Dibble, H. L. (2018). Experimental design and experimental inference in stone artifact archeology. *Journal of Archeological Method and Theory*, 25(3), 663–688. <https://doi.org/10.1007/s10816-017-9351-1>.
- Lin, S. (2014). Experimentation and scientific inference building in the study of hominin behavior through stone artifact archeology. Publicly Accessible Penn Dissertations.
- Longacre, W. (2010). Archeology as anthropology revisited. In *Journal of Archeological Method and Theory* (Vol. 17, pp. 81–100). UNP - Nebraska. doi:<https://doi.org/10.2307/40784755>.
- Longo, L., & Skakun, N. (2008). Prehistoric technology' 40 years later: functional studies and the Russian legacy. Proceedings of the international congress Verona (Italy), 20–23 April 2005.
- Macdonald, D. A., Harman, R., & Evans, A. A. (2018). Replicating surface texture: preliminary testing of molding compound accuracy for surface measurements. *Journal of Archeological Science: Reports*, 18, 839–846. <https://doi.org/10.1016/j.jasrep.2018.02.033>.
- Mackay, A., Lin, S., Kenna, L., & Blackwood, A. (2018). Variance in the response of silcrete to rapid heating complicates assumptions about past heat treatment methods. *Archeological and Anthropological Sciences*. <https://doi.org/10.1007/s12520-018-0663-1>.
- Magnani, M., Rezek, Z., Lin, S. C., et al. (2014). Flake variation in relation to the application of force. *Journal of Archeological Science*, 46, 37–49. <https://doi.org/10.1016/j.jas.2014.02.029>.
- Marean, C., Bar-Matthews, M., Bernatchez, J., Fisher, E., Goldberg, P., Herries, A., et al. (2007). *Early human use of marine resources and pigment in South Africa during the Middle Pleistocene* (Vol. 449). <https://doi.org/10.1038/nature06204>.
- Marreiros, J., Mazzucco, N., Gibaja, J., & Bicho, N. (2015). Macro and micro evidences from the past: the state of the art of archeological use-wear studies. In J. M. Marreiros, J. F. Gibaja Bao, & N. Ferreira Bicho (Eds.), *Use-wear and residue analysis in archeology* (pp. 5–26). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-08257-8_2.
- Martisius, N., Sidéra, I., Grote, M., Steele, T., McPherron, S., & Schulz-Komas, E. (2018). Time wears on: assessing how bone wears using 3D surface texture analysis. *PLoS One*, 13(11), e0206078. <https://doi.org/10.1371/journal.pone.0206078>.
- McBrearty, S., & Brooks, A. S. (2000). The revolution that wasn't: a new interpretation of the origin of modern human behavior. *Journal of Human Evolution*, 39(5), 453–563. <https://doi.org/10.1006/jhev.2000.0435>.
- McCall, G. (2012). Ethnoarcheology and the organization of lithic technology. *Journal of Archeological Research*, 20(2), 157–203. <https://doi.org/10.1007/s10814-011-9056-z>.
- McPherron, S., Braun, D., Dogandžić, T., Archer, W., Desta, D., & Lin, S. (2014). An experimental assessment of the influences on edge damage to lithic artifacts: a consideration of edge angle, substrate grain size, raw material properties, and exposed face. *Journal of Archeological Science*, 49, 70–82. <https://doi.org/10.1016/j.jas.2014.04.003>.
- Miller, R., & Hayden, B. (2006). Palaeolithic reflections: lithic technology and ethnographic excavations among Australian aborigines. *Man*, 17(3), 561. <https://doi.org/10.2307/2801725>.
- Morgan, T., Uomini, N., Rendell, L., Chouinard-Thuly, L., Street, S., Lewis, H., et al. (2015). Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature Communications*, 6(1), 6029. <https://doi.org/10.1038/ncomms7029>.
- Moss, E.H., 1987. A review of "investigating microwear polishes with blind tests". *Journal of Archeological Science* 14, 473–481.
- Nance, J. (1994). Statistical sampling, estimation, and analytic procedures in archeology. *Journal of Quantitative Anthropology*, 4(3), 221–248.
- Newcomer, M. H., Grace, R., & Unger-Hamilton, R. (1986). Investigating microwear polishes with blind tests. *Journal of Archeological Science*, 13, 203–217.
- Newcomer, M. H., Grace, R., & Unger-Hamilton, R. (1988). Microwear methodology: a reply to Moss. *Hurcombe and Bamforth. Journal of Archeological Science.*, 15, 25–33.
- Nonaka, T., Bril, B., & Rein, R. (2010). How do stone knappers predict and control the outcome of flaking? Implications for understanding early stone tool technology. *Journal of Human Evolution*, 59(2), 155–167. <https://doi.org/10.1016/j.jhev.2010.04.006>.
- Odell, G. (2001). Stone tool research at the end of the millennium: classification, function, and behavior. *Journal of Archeological Research*, 9(1), 45–100. <https://doi.org/10.1023/A:1009445104085>.
- Odell, G. (2000). *Stone tool research at the end of the millennium: procurement and technology* (Vol. 8). <https://doi.org/10.1023/A:1009439725979>.
- Odell, G. (1981). The mechanics of use-breakage of stone tools: some testable hypotheses. *Journal of Field Archeology*, 8(2), 197–210.

- Odell, G. (1975). Micro-wear in perspective: a sympathetic response to Lawrence H. Keeley. *World Archeology*, 7(2), 226–240. <https://doi.org/10.1080/00438243.1975.9979635>.
- Ollé, A., & Vergès, J. (2014). The use of sequential experiments and SEM in documenting stone tool microwear. *Journal of Archeological Science*, 48(1), 60–72. <https://doi.org/10.1016/j.jas.2013.10.028>.
- Outram, A. (2008). Introduction to experimental archeology. *World Archeology*, 40(1), 1–6. <https://doi.org/10.1080/00438240801889456>.
- Pedergrana, A. (2017). Microwear and residue analyses of quartzite stone tools. Experimental development of a method and its application to the assemblages from the Pleistocene sites of Gran Dolina-TD10 (Sierra de Atapuerca, Burgos, Spain) and Payre (Ardèche, France). Phd Thesis submitted to the University of Tarragona (Spain) and the Muséum National d'Histoire Naturelle de Paris (France).
- Pedergrana, A., & Ollé, A. (2017). Monitoring and interpreting the use-wear formation processes on quartzite flakes through sequential experiments. *Quaternary International*, 427, 35–65. <https://doi.org/10.1016/j.quaint.2016.01.053>.
- Pedergrana, A. (2019). “All that glitters is not gold”: evaluating the nature of the relationship between archeological residues and stone tool function. *J Paleo Arch*. <https://doi.org/10.1007/s41982-019-00039-z>.
- Pfleging, J., Stücheli, M., Iovita, R., & Buchli, J. (2015). Dynamic monitoring reveals motor task characteristics in prehistoric technical gestures. *PLoS One*, 10(8), e0134570. <https://doi.org/10.1371/journal.pone.0134570>.
- Pfleging, J., Iovita, R., & Buchli, J. (2019). Influence of force and duration on stone tool wear: Results from experiments with a force-controlled robot. *Archaeological and Anthropological Sciences*, 11(11), 5921–5935.
- Phillips, P. (1988). Traceology (microwear) studies in the USSR. *World Archeology*, 19(3), 349–356. <https://doi.org/10.1080/00438243.1988.9980045>.
- Piggott, S., & Trigger, B. G. (1991). A history of archeological thought. *Journal of Field Archeology*, 18(1), 109. <https://doi.org/10.2307/530155>.
- Rots, V., Hardy, B., Serangeli, J., & Conard, N. (2015). Residue and microwear analyses of the stone artifacts from Schöningen. *Journal of Human Evolution*, 89, 298–308. <https://doi.org/10.1016/j.jhevol.2015.07.005>.
- Rots, V., & Williamson, B. S. (2004). Microwear and residue analyses in perspective: the contribution of ethnoarchaeological evidence. *Journal of Archeological Science*, 31, 1287–1299. <https://doi.org/10.1016/j.jas.2004.02.009>.
- Rots, V., Hayes, E., Cnats, D., Lepers, C., & Fullagar, R. (2016). Making sense of residues on flaked stone artefacts: learning from blind tests. *PLoS One*, 11(3), e0150437. <https://doi.org/10.1371/journal.pone.0150437>.
- Rots, V., & Plisson, H. (2014). *Projectiles and the abuse of the use-wear method in a search for impact* (Vol. 48). doi:<https://doi.org/10.1016/j.jas.2013.10.027>.
- Rots, V. (2013). *Prehension and hafting traces on flint tools*. Leuven University Press. doi:<https://doi.org/10.11116/9789461660060>.
- Sahle, Y., Hutchings, W. K., Braun, D., Sealy, J. C., Morgan, L., Negash, A., & Atnafu, B. (2013). Earliest stone-tipped projectiles from the Ethiopian rift date to >279,000 years ago. *PLoS One*, 8(11). <https://doi.org/10.1371/journal.pone.0078092>.
- Sanchez-Yustos, P., Diez-Martin, F., Dominguez-Rodrigo, M., Duque, J., Fraile, C., Diaz, I., et al. (2017). The origin of the Acheulean. Techno-functional study of the FLK W lithic record (Olduvai, Tanzania). *PLoS One*, 12(8):e0179212.
- Schiffer, M. (1979). The place of lithic use-wear studies in behavioural archeology. In *Lithic use-wear analysis* (pp. 15–25). New York, USA: Academic Press.
- Schmidt, P., & Mackay, A. (2016). Why was silcrete heat-treated in the middle stone age? An early transformative technology in the context of raw material use at mertenhof rock shelter, South Africa. *PLoS One*, 11, e0149243. <https://doi.org/10.1371/journal.pone.0149243>.
- Schmidt, P., Blessing, M., Rageot, M., et al. (2019). Birch tar production does not prove Neanderthal behavioral complexity. *PNAS*, 201911137. <https://doi.org/10.1073/pnas.1911137116>.
- Schmitt, D., Churchill, S., & Hylander, W. (2003). Experimental evidence concerning spear use in Neandertals and early modern humans. *Journal of Archeological Science*, 30(1), 103–114. <https://doi.org/10.1006/jasc.2001.0814>.
- Schoville, B.. (2014). Testing a taphonomic predictive model of edge damage formation with Middle Stone Age points from Pinnacle Point Cave 13B and Die Kelders Cave 1, South Africa. *Journal of Archeological Science*, 48, 84–95. doi:<https://doi.org/10.1016/j.jas.2013.10.002>.

- Schoville, B., Wilkins, J., Ritzman, T., Oestmo, S., & Brown, K. (2017). The performance of heat-treated silcrete backed pieces in actualistic and controlled complex projectile experiments. *Journal of Archeological Science: Reports*, 14, 302–317. <https://doi.org/10.1016/j.jasrep.2017.05.053>.
- Scott, R., Ungar, P., Bergstrom, T., Brown, C. a, Grine, F., Teaford, M., & Walker, A. (2005). Dental microwear texture analysis shows within-species diet variability in fossil hominins. *Nature*, 436(7051), 693–695. doi:<https://doi.org/10.1038/nature03822>.
- Semenov, S. (1957). *Pervobytnaja tehnika*. Moskva-Leningrad: Nauka.
- Semenov, S. A. (1964). Prehistoric technology: an experimental study of the oldest tools and artefacts from traces of manufacture and wear. Cory, Adams and Mackay, London.
- Semenov, S. (1970). Forms and functions of the oldest stone tools. *Quartär*, 21, 1–20.
- Shea, J. (1987). On accuracy and relevance in lithic use-wear analysis. *Lithic Technology*, 16(2/3), 44–50.
- Shea, J. (2011). Stone tool analysis and human origins research: some advice from uncle screwtape. *Evolutionary Anthropology: Issues, News, and Reviews*, 20(2), 48–53. <https://doi.org/10.1002/evan.20290>.
- Stemp, W. (2014). A review of quantification of lithic use-wear using laser profilometry: a method based on metrology and fractal analysis. *Journal of Archeological Science*, 48, 15–25. <https://doi.org/10.1016/j.jas.2013.04.027>.
- Stemp, W., & Chung, S. (2011). Discrimination of surface wear on obsidian tools using LSCM and ReLA: pilot study results (area-scale analysis of obsidian tool surfaces). *Scanning*, 33(5), 279–293. <https://doi.org/10.1002/sca.20250>.
- Stemp, W., Lerner, H., & Kristant, E. (2013). Quantifying microwear on experimental Mistassini quartzite scrapers: preliminary results of exploratory research using LSCM and scale-sensitive fractal analysis. *Scanning*, 35(1), 28–39. <https://doi.org/10.1002/sca.21032>.
- Stemp, W., Lerner, H., & Kristant, E. (2018). Testing area-scale fractal complexity (Asfc) and laser scanning confocal microscopy (LSCM) to document and discriminate microwear on experimental quartzite scrapers. *Archaeometry*, 60(4), 660–677. <https://doi.org/10.1111/arc.12335>.
- Stemp, W., Macdonald, D. A., & Gleason, M. A. (2019). Testing imaging confocal microscopy, laser scanning confocal microscopy, and focus variation microscopy for microscale measurement of edge cross-sections and calculation of edge curvature on stone tools: preliminary results. *Journal of Archeological Science: Reports*, 24, 513–525. <https://doi.org/10.1016/j.jasrep.2019.02.010>.
- Stemp, W., Morozov, M., & Key, A. J. M. (2015). Quantifying lithic microwear with load variation on experimental basalt flakes using LSCM and area-scale fractal complexity (Asfc). *Surface Topography: Metrology and Properties*, 3(3), 034006. <https://doi.org/10.1088/2051-672X/3/3/034006>.
- Stemp, W., & Stemp, M. (2001). UBM laser profilometry and lithic use-wear analysis: a variable length scale investigation of surface topography. *Journal of Archeological Science*, 28(1), 81–88. <https://doi.org/10.1006/jasc.2000.0547>.
- Stemp, W., & Stemp, M. (2003). Documenting stages of polish development on experimental stone tools: surface characterization by fractal geometry using UBM laser profilometry. *Journal of Archeological Science*, 30(3), 287–296. <https://doi.org/10.1006/jasc.2002.0837>.
- Stemp, W., Watson, A., & Evans, A. (2016). Surface analysis of stone and bone tools. *Surface Topography: Metrology and Properties*, 4(1), 013001. <https://doi.org/10.1088/2051-672X/4/1/013001>.
- Stout, D., Passingham, R., Frith, C., Apel, J., & Chaminade, T. (2011). Technology, expertise and social cognition in human evolution. *European Journal of Neuroscience*, 33(7), 1328–1338. <https://doi.org/10.1111/j.1460-9568.2011.07619.x>.
- Tomenchuk J. (1985). The development of a wholly parametric use-wear methodology and its application to two selected samples of epipaleolithic chipped stone tools from hayonim cave, Israel [microform]. Thesis (Ph.D.)-University of Toronto.
- Trigger, B. G. (1989). A history of archaeological thought. Cambridge University Press.
- Tringham, R., Cooper, G., Odell, G., Voytek, B., & Whitman, A. (1974). Experimentation in the formation of edge damage: a new approach to lithic analysis. *Journal of Field Archeology*, 1(1–2), 171–196. <https://doi.org/10.1179/jfa.1974.1.1-2.171>.
- Tumung, L., Bazgir, B., & Ollé, A. (2015). Applying SEM to the study of use-wear on unmodified shell tools: an experimental approach. *Journal of Archeological Science*, 59(May), 179–196. <https://doi.org/10.1016/j.jas.2015.04.017>.
- Ungar, P., & Evans, A. (2016). Exposing the past: surface topography and texture of paleontological and archeological remains. *Surface Topography: Metrology and Properties*, 4(4), 040302. <https://doi.org/10.1088/2051-672X/4/4/040302>.
- van Gijn, A. (2014). Science and interpretation in microwear studies. *Journal of Archeological Science*, 48(1), 166–169. <https://doi.org/10.1016/j.jas.2013.10.024>.

- Williams-Hatala, E., Hatala, K., Gordon, M., Key, A., Kasper, M., & Kivell, T. (2018). The manual pressures of stone tool behaviors and their implications for the evolution of the human hand. *Journal of Human Evolution*, 119, 14–26. <https://doi.org/10.1016/j.jhevol.2018.02.008>.
- Zupancich, A., Nunziantese-Cesaro, S., Blasco, R., Rosell, J., Cristiani, E., Venditti, F., et al. (2016). Early evidence of stone tool use in bone working activities at Qesem cave, Israel. *Scientific Reports*, 6, 37686.

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

Affiliations

João Marreiros^{1,2,3} · Ivan Calandra¹ · Walter Gneisinger¹ · Eduardo Paixão^{1,2,3} · Antonella Pedergnana¹ · Lisa Schunk^{1,2}

¹ TraCER, Laboratory for Traceology and Controlled Experiments at MONREPOS Archaeological Research Centre and Museum for Human Behavioural Evolution, RGZM, Schloss Monrepos, 56567 Neuwied, Germany

² Institute for Prehistoric and Protohistoric Archeology, Johannes Gutenberg University, Schönborner Hof, Schillerstraße 11, 55116 Mainz, Germany

³ ICArEHB, Interdisciplinary Center for Archaeology and Evolution Human Behaviour, University of Algarve, Campus de Gambelas, 8005-139 Faro, Portugal