



Universiteit
Leiden
The Netherlands

Reconstruct the unknown, replicate the uncontrollable: current issues in the experimental archaeology of combat

Gentile, V.

Citation

Gentile, V. (2022). Reconstruct the unknown, replicate the uncontrollable: current issues in the experimental archaeology of combat. *Journal Of Archaeological Science: Reports*, 46. doi:10.1016/j.jasrep.2022.103709

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](#)

Downloaded from: <https://hdl.handle.net/1887/3561248>

Note: To cite this publication please use the final published version (if applicable).



Reconstruct the unknown, replicate the uncontrollable. Current issues in the experimental archaeology of combat

Valerio Gentile

Faculty of Archaeology, Leiden University, the Netherlands

ARTICLE INFO

Keywords:

Experimental archaeology
Metalwork wear analysis
Combat experiments
Bronze Age
Weapons

ABSTRACT

Following the so-called material turn, in archaeology much attention is devoted to the affective response to objects, the physical affordances of items, or the agency of materials on one another. However, such aspects have been partially overlooked in experimental use wear research. Issues surrounding contact material selection and its degree of representativity against the larger archaeological sample are some of the problems that a well-considered approach in experimental archaeology and wear analysis should take into account. These problems are inherently linked with the discussion over controlled lab experiments vs actualistic layouts: one of the most contentious debates in experimental archaeology. More broadly, these issues are further tied to the crisis of confidence in experimental results and issues such as replicability and reproducibility. These concerns are even more significant in research attempting to simulate and investigate combat wear traces, where these problems also intertwine with the challenges that these layouts pose in terms of best practices to follow to ensure ethics and sustainability. In this paper, the methodological framework implemented in two experimental campaigns studying prehistoric bronze weaponry is discussed. The examples are then used to illustrate some of the challenges in these types of set-ups and to provide discussion points regarding potential solutions. In addition, steps to take in order to increase confidence in the interpretation of experimental results are proposed. While replication of experimental results is paramount, it is also necessary to reduce the ambiguity of experimental results.

1. Introduction:

The most recurrent unit of analysis for the interaction between a human agent and an item in archaeology is probably represented by the relationship between the user and their tool(s). Some of the most reliable methods in the archaeological toolbox to gain insights and test assumptions on what such interactions entailed in the past are (micro) wear analysis paired with experimental archaeology. The tool acts as an intermediary between the human body and other matter. In other words, what comes to be is a tool-mediated dialogue between a user and an external entity. It is contact with matter that makes objects develop wear traces that archaeologists can then study to reconstruct the interactions that caused them. What follows is that, together with the archaeological item, 'contact material' should play a critical role in both theoretical considerations around materiality, as well as, more practically, in the interpretation of wear patterns and in the design of experimental trials.

A considered approach to experimental archaeology and wear analysis passes necessarily through discussions over issues surrounding the choice of contact material and its degree of representativity of the

archaeological sample. Such matters are tied to one of the most important discussions in experimental archaeology: the 'control vs actualism' debate, which is – in turn – further connected to the modern crisis of confidence towards experimental results and issues such as the replicability and reproducibility of research and results (cf. Baker 2016). Finally, discussion on the validity of contact materials would benefit from engaging critically in issues such as the best practices to follow to safeguard ethics and sustainability.

These matters are even more pressing for relatively recent and complex branches of experimental archaeology such as combat experiments with metal weaponry. In this paper, I discuss two experimental campaigns with bronze weaponry to highlight some of the challenges encountered and present possible approaches to adopt in order to mitigate the issues reviewed. Additionally, I suggest further steps to take to increase confidence in the interpretation based on experimental results.

E-mail address: v.gentile@arch.leidenuniv.nl.

<https://doi.org/10.1016/j.jasrep.2022.103709>

Received 30 April 2022; Received in revised form 15 October 2022; Accepted 23 October 2022

Available online 10 November 2022

2352-409X/© 2022 The Author. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

2. The tension between analogy and control

Within the discipline of experimental archaeology, the degree to which variable control should be exerted and its influence on replicability and generalisation, has been a subject of long debate. One of the most important points of discussion has been the tension between designing tests that grant extensive control and replicability and creating experiments that more closely match the phenomena under study. Callahan (1999) famously operated a sharp distinction between valid and invalid experiments in archaeology: to be valid, an experiment must be designed in such a way that it can be replicated in the future, and constant attention to recording is required at all stages. Notably, he also remarked that in case a specific human action is part of the test, the performance must not be influenced by the learning of the skills by the performer (*ibid.*, 5).

In principle, the more variables are under control, the more it is possible to obtain quantifiable results that can be replicated and tested by other researchers. As a result, laboratory tests that utilize controlled environments and mechanized facilities are popular in the discipline. However, a gap exists between the laboratory experiment and how the same action/processes took place in the past, with limited technology, little control over the environment, and increased variability (Outram 2008). Another possible route is, then, to attempt to recreate stronger analogy with what might have happened in the past in 'actualistic' experiments. Self-evidently, in the interest of representativeness, these trials forfeit a significant degree of control (Schenck 2011, 87-8).

Rather than being radically opposed, the two approaches can be considered parts of the same continuum (cf. Dolfini and Collins 2018, Mathieu 2002). Highly controlled experiments, on the one hand, are useful for discriminating between dependent and independent variables and hence for testing hypotheses. However, while the controlled nature of such studies makes them – in theory – easier to replicate, it also makes them less imitative of 'regular' human activities. Experiments with less control over the variables, on the other hand, are generally less suitable for distinguishing between dependent and independent variables, as well as testing specific hypotheses. They do, however, tend to imitate more effectively past human behaviours and are an invaluable tool to generate inferences.

Experiments with metal implements and more specifically combat experiments are particularly suitable case studies to consider for several reasons. Due to its novelty (see below), the field is still in the process of reaching methodological maturity and forming shared protocols (Dolfini and Crellin 2016). Furthermore, if this type of activity is problematic to recreate faithfully in a simplified lab-environment, at the same time, the variety and pace of actualist settings pose serious challenges for control and replicability. Finally, the nature of the tests requires important consideration in matters such as safety and ethics. A succinct overview of how the discipline developed over time is provided below before discussing the methodological layout of two experimental campaigns focusing on Bronze Age weaponry in detail.

3. Combat experiments with bronze weaponry

Experimental archaeology and wear analysis on bronze implements and weapons has only recently left its infancy and it is gradually reaching methodological maturity (Dolfini and Crellin 2016). For what concerns combat experiments, the discipline has developed over time from highly controlled lab experiments to a more experiential approach, to recent efforts to strike a balance between an acceptable amount of control while maintaining a high degree of analogy (see Crellin et al. 2018 for an extensive critical review). The work of Bridgford (2000) represents the first lab-based approach to the reconstruction of combat traces on Bronze Age weaponry. In her set-up, segments of bronze intended to simulate swords' edges were attached to a mechanized rig and made to collide against other edge-segments held statically. A similar approach was implemented by O'Flaherty and colleagues (2011)

for the reproduction of combat damage on replicas of Early Bronze Age halberds utilizing a Rosand machine. Conversely, Molloy undertook an experiential appraisal of Bronze Age weaponry (e.g. 2008, 2010). His research was not explicitly designed to reproduce and document wear traces but mainly concerned with the assessment of the functionality of Bronze Age swords which he investigated by performing cut-test mats and pig legs. Although those trials retained less control over the variables involved, they demonstrated the degree of effectiveness of pre-historic weaponry and, importantly, showed how different ways of performing an attack (informed by combat training) significantly affect the offensive capabilities of the weapons; a circumstance that controlled experiments were not able to reproduce or observe. Anderson (2011) designed and carried out experiments aimed at replicating combat traces on spears through actualistic experiments with the help of martial arts trained volunteers. Although sparse information over variable control and gaps in documentation undermine the replicability of the study, this experiment had a considerable impact on the direction undertaken by later research on wear formation on weaponry. Recently, a new strand of experimental studies devoted to the understanding of wear formation on prehistoric weaponry has emerged (e.g. Gentile and van Gijn, 2019, Gentile et al. in prep, Hermann et al. 2020a, Knight 2019), in which tests are consciously designed and carried out so as to guarantee that a sufficient amount of key variables can still be monitored and documented whilst attempting to maintain enough analogy with the past action.

Such a hybrid approach stems from the awareness that in combat situations, as weapons clang against their homologue, they are both held by mirroring agents, the combatants, who continuously react to each other's gestures (Gentile and van Gijn 2019, Hermann et al. 2020b). Although enough control over the main variables involved and thorough documentation are paramount, such characteristics make it challenging not to implement actualistic frameworks and not to resort to human participants.

3.1. Controlling mayhem – An attempt to conciliate actualism and variable control in combat experiments

In this section, I will discuss the strategies implemented to conciliate actualism and variable control in two experimental campaigns focusing on Bronze Age combat and weaponry (Gentile and van Gijn 2019, Gentile et al. in prep; Fig. 1). With the aim of assessing the use in combat of Late Bronze Age/Early Iron Age swords, van Gijn and I teamed up with expert practitioners of ancient fencing and historical martial arts to conceive and perform a series of controlled but realistic tests (Gentile and van Gijn 2019). The main goals of the experiments were to understand to what extent combat produced traces similar to those sometimes identified on archaeological swords, and to what extent different combat movements resulted in different traces.

The tests were designed to reproduce a series of collisions which were deemed to be the most elementary units of combat possible to perform with such weapons, using historical fencing manuals as a conceptual biomechanical scaffolding but devoid of any school-specific precept (on the challenges of reconstructing Bronze Age combat techniques with the help of historical combat manuals see Gentile and van Gijn 2019, and Hermann et al. 2020b). The routine was broken down into single, synchronous movements of attack and defences to ensure control over the action and enable documentation after each collision. At the same time this approach allowed the collisions to stay analogous to a real combat scenario. For the sake of repeatability and in order to assess the degree of uniformity of the results, each combination was repeated a minimum of two times, with each expert taking turns attacking and defending. The wear traces produced after each collision were then documented and studied under the microscope.

These tests showed promising results, casting additional light on the formation dynamics of wear traces on bronze swords. Furthermore, the results suggested that the frequency of development of certain traces is likely correlated with the combat style implemented. A comparison with



Fig. 1. Some of the replicas used in the experimental campaigns discussed. Image: V. Gentile

wear found on a sample of archaeological swords further validated the results (Gentile and van Gijn 2019). Once the methodology was successfully tested, a wider, multi-stage, experimental campaign focusing on bronze spears was performed in 2020 in collaboration with researchers and practitioners of Historical European Martial Arts (Gentile et al. in prep, van Dijk 2020). The new campaign capitalized on the double nature of experimental archaeology as hypothesis-generating as well as hypothesis-testing discipline (Dolfini and Collins 2018, Lammers-Keijsers 2005) by creating a series of experiments in which the results of a test are at the same time informative for, and further assessed by, the performance and outcome of the next test. The result is a workflow in which the ratio between control and actualism is gradually shifted in order to strengthen interpretation and increase generalisation.

The first experiment followed very closely the replicable methodology of the sword combat. In these first trials, a relatively high level of control was exerted by having expert users make spears collide multiple times, in specific areas, at specific angles. In addition to testing the correlation between specific combat combinations and the production of distinct traces, these tests were intended to gain further insight into the degree of variability in trace formation, as well as into the material integrity and durability of the weapons.

In the second experiment, strikes against an animal carcass were performed. Together with generating data on the wear that skin and bone could produce on bronze weaponry, assumptions about what kind of strength and movements were necessary to inflict specific kinds of trauma (e.g. deep/lethal vs superficial/non-lethal wounds) were also tested. This experiment further envisaged two shifting levels of control: one phase in which specific strikes were performed separately (similarly to experiment one) and another phase in which a series of attacks was performed in a chained motion analogous to a real combat situation.

In the third experiment, the level of analogy was pushed to the extreme to confirm, refine, and expand the insights gained in the

previous experiments. Contrary to what performed before, the expert users were not tasked with reproducing specific movements but with sparring freely according to certain pre-established combat styles (Fig. 2). Action was only suspended when a natural break in combat occurred. Such a multi-stage approach allowed the assessment of to what extent results of more controlled experiments remain coherent when actualism increases and other less controlled variables (e.g. varying distance and opportunity) are introduced. Furthermore, gradually introducing less control in the set-up allowed the observation of phenomena which could not have been identified maintaining the original set of conditions.

Gradually shifting from more controlled to less controlled setups can increase confidence in the interpretation of the results and their potential for generalisation. Nevertheless, the pursuit of stronger analogy poses several challenges, many of these being directly connected to the choice of contact material. In the next section, some of the main challenges encountered will be discussed together with the workarounds implemented.

3.2. Challenges and compromises in actualistic combat experiments

Performing actualistic combat experiments while maintaining sufficient control requires 'locking' at least some of the main variables involved. This calls for several crucial decisions aimed at striking a balance between replicability and potential for analogy to a less uniform archaeological sample. Moreover, due to the very nature of the tests, challenges to both replicability and genuine analogy come also from the areas of ethics or safety.

In certain combat tests not only the combatants' choices matter but a third human actor is always present and playing a decisive role: namely the craftsman, present in the form of the physical properties of the weaponry. Alloy composition, casting techniques, and post casting



Fig. 2. A moment of the third spear combat experiment, with combatants sparring freely. Image: V. Gentile

treatments (e.g. work hardening) can all alter the characteristics of the item and the wear traces that it develops (Soriano-Llopis and Gutiérrez-Sáez, 2009). This is even more relevant in tests involving metal weapons (e.g. two swords colliding), as the dichotomy between tool and contact material shifts from absolute to relative: each implement is at the same time regarded as the surface developing as well as causing the trace. Such variables can be efficiently kept under control by operating collisions between implements with the same characteristics. Nevertheless, by doing so, one forfeits – for the sake of control and replicability – a certain degree of analogy. In fact, it is unlikely that, in the past, weapons collided only against items with exactly the same properties.

Besides for composition, Bronze Age metal items also greatly differed according to the crafting techniques employed and skill level of the artisan (Kuijpers, 2017). Nowadays, techniques for crafting bronze implements are being independently reverse engineered by a small number of craftspeople distributed across the globe, who follow different traditions and protocols. The scant availability of experienced craftspeople, together with the small number of experiments so far conducted, results in low representativity when it comes to the items produced: for instance, the vast majority of experimental archaeology concerning bronze implements -and weapons in particular- performed until now have featured replicas crafted by a single craftspeople which likely sits in the most skilled side of the spectrum (cf. Anderson 2011, Downing and Fibiger 2017, Hermann et al. 2020a, Knight 2019, Molloy 2008). If on one hand these limitations represent a welcome scenario, in which replicability is thoroughly ensured, on the other hand one is compelled to consider to what extent this would affect the possibility to draw general conclusions on the behaviour the Bronze Age weapons found across Europe. In order to mitigate this effect, besides wishing for a general increase in the number of tests, a feasible option could be to keep uniformity of characteristics within the experiments but vary composition and crafting techniques across experiments. Nonetheless, in the long run, controlled experiments specifically designed to assess the impact of each crafting choice on trace formation would be needed to completely dispel ambiguity (see also section 4 below).

In order to avoid injury or death in combat, protective implements are generally opposed to weapons. Protective gear thus represents another important contact material to consider and investigate. In combat, fighters defend from the opponents' attacks by using their own weapons or specific implements to deflect or absorb the blow. Besides testing blade vs blade contact, some of the spear combat tests described above (and in Gentile et al. in prep), envisaged the use of wooden shields (Figs. 1 and 2). Such a decision was operated with two main goals in mind: 1 – to provide the combatants with a homologue of a Bronze Age shield to better inform their gestures and present them with analogous restrictions, 2 – to provide a contact surface plausible enough to generate insights on the wear that weapons could develop against generic Bronze Age shields.

Although several metal shields are known (Molloy 2009, Uckelmann 2012), the vast majority of Bronze Age shields were likely made of perishable materials. Unfortunately, only very few specimens of shields made of wood or leather have survived (ibid). The Bronze Age wooden shield found in Annandale (Ireland), was taken as a general archetype, and replicas have reproduced its measurements taking the shrinkage of archaeological wood into account.

Nevertheless, replicating the same material properties of the original shield is problematic. The Annandale shield was made from a single piece of Alder (*Alnus*). Unfortunately, modern wood-cutting practices make the acquisition of a piece of Alder suitable for replicating the shield quite challenging. On the other hand, the Annandale specimen represents a unicum in the archaeological record. Therefore, testing other woods as contact material has the advantage of fine-tuning generalizations on the wear that perishable material shields could have left on ancient bronze weaponry. For these reasons, much more available spruce wood (*Picea abies*) was used to craft the shield homologues.

Changes in forestry procedures as well as environmental laws can pose limitations that directly affect the degree of analogy that can be achieved in reproducing specific items. In cases like these, it also directly affects the possibility for different research teams to replicate the experiments on the basis, for example, of local legislations and policies. On

the other hand, as previously discussed in the context of weapon properties, chasing absolute replication can affect generalization.

Regardless of the degree of actualism one aims to implement in their experimental layouts, certain activities are practically impossible to reproduce. Inflicting combat wounds to another living – human – being is perhaps the clearest example of these. In previous experiments aiming at simulating contact between bronze weapons and a human body, blows were landed against animal carcasses (Anderson 2011, Molloy 2008, O’Flaherty 2007) or synthetic bone material (Downing and Fibiger 2017).

The second experiment of the spear combat tests (Gentile et al. in prep, van Dijk 2020) made use of animal contact material. This choice was operated on the basis of the experiment’s objective (see section 3.1), and to generate more comparable results to the majority of previous research. Nevertheless, this choice also comes with some challenges: matters such as opportunity, logistics, as well as ethics all play a role. In order not to commission any animal killing, it was decided to contact a specialized butcher for game and acquire the first available medium-size carcass. This meant that the animal could not be actively selected, that the time window to prepare and enact the tests would have been restricted, and –most importantly– that all the muscle tissues and interior organs would have been removed by the butcher leaving only skin and bone material.

The first carcass available was one of a young roe deer (*Capreolus capreolus*). The removal of muscles mass is expected to affect the degree of analogy for certain areas of the human body, such as the abdomen. However, direct attachment of skin to bone made the carcass better resemble areas of the body commonly targeted during combat, such as the head, ribcage, and forearms, where the skin and bone are not separated by thick layers of muscle. Human skin thickness varies considerably but it is on average around 1.2 mm (Lee and Hwang 2002), while roe deer skin in females and subadults stays below 2 mm (Sokolov and Danilkin 1979). In the specimen used in the experiments, the skin surrounding the areas hit was c.1 mm thick. Considering the aims of the tests, the analogy was deemed satisfactory. To better resemble human skin and facilitate the identification of the wounds, the deer skin was dehaired. The carcass was hung from a suspension system allowing the target to move slightly when hit, while offering a resistance similar to the resistance a standing human would offer.

Although the compromises to be made in acquiring and using animal material undoubtedly limit the type of research questions that could be answered, they also bring benefits in terms of ethics and sustainability. In the experiment here described, it was made use of what would have been otherwise considered ‘waste material’ and it was not, in any way, commissioned or economically rewarded any animal killing. Evidently, local laws and guidelines from ethical committees could influence the carrying out of experiments on animal materials (rightfully so), directly affecting the replicability of certain tests. The compromises described above are also expected to enable a decent degree of replicability, as it is more ethical, accessible, and sustainable to set up experiments which use already sourced material, while maintaining some degree of comparability with past tests on animal carcasses. Nevertheless, with the growing progress in the manufacturing and accessibility of synthetic analogues of human tissues, it is desirable that, in the future, experiments will increasingly rely on artificial material.

Finally, despite all the control efforts, the human factor of the actualistic setups is bound to represent a variable equally impactful as it is hard to control. In order to guarantee as much replicability as possible in actualistic experiments it is pivotal that the – combat – movements performed are properly described. Nevertheless, it is known that written text is not the most efficient way to transmit what is often learned by doing. One can and to some extent should compensate this through a thorough video and photographic documentation (cf. Gentile and van Gijn 2019, Hermann et al. 2020a). However, it is worth considering to what extent, given the inductive nature of investigating long-gone combat techniques, a faithful but passive imitation of previous tests is

entirely beneficial. Rather than replicating accurately previously tested combat combinations, experimenters and expert users could take into account only a few important parameters and reproduce analogous situations according to their own style and bodily knowledge. Aiming at reproduction rather than complete replication would avoid interfering with Callahan’s precept of not making the process of learning a new a skill hamper the test (1999). Furthermore, it would introduce small variable changes instrumental in achieving a better generalisation (see section 4 below).

On the ethical side, having human participants involved in the experiments comes with important considerations regarding safety. Making sure to collaborate with expert practitioners, besides respecting Callahan’s recommendation of keeping learning out the experiment (see above), also grants an essential layer of safety. The researcher should make sure that the expert practitioner is in the position of advising and influencing the experimental layout in order to make it as safe as possible, due to their deeper familiarity with the tools and movements involved. In the experimental campaigns discussed in this article, the dialogue with experts was instrumental to the creation of the proper layout and in adopting all the necessary safety measures. To some extent, it can be argued that wearing modern protective gear could influence the gesture and range of movements available to the practitioners as well as altering their confidence and approach, ultimately at the expenses of analogy towards the situation one wants to replicate (Jaquet et al. 2015). Although this is plausible, the possibility that such discrepancies could play a substantial role in the wear formation processes remains to be demonstrated (provided that contact with modern protective gear should be avoided and thoroughly documented should it occur). Finally, similarly to the issues concerning the use of animal and plant material, the performance of combat tests in the most analogous way as possible to the real scenario, might also pose replicability issues. Elements such as research and liability regulations might not allow the exact replication of a specific layout across different countries and institutions.

4. Moving forward. Triangulation and tackling ambiguity

So far, the challenges of designing and performing experiments that enable a decent amount of realism without compromising replicability and reproducibility have been discussed together with strategies and workarounds implemented to obtain optimal results. These are only some of the problems which can affect the performance and replication of the experiments. Issues such as the limited number of repetitions performed in combat experiments (often conditioned by the high cost of the replicas), the absence of a shared nomenclature for the description of the traces, and the inherent difficulty in properly documenting user movements in actualistic scenarios are important issues whose in-depth discussion goes beyond the scope of this paper, which is mainly concerned with issues surrounding contact material.

Undoubtedly, awareness of the limitations of the discipline is paramount to improve our research and interpretations, especially in a phase in which replicability issues and the partially consequent crisis in the confidence towards experimental conclusions have reached the field of experimental archaeology. This is even more relevant when it comes to such a young and complex sub-field such as combat tests with (metal) weaponry. Nevertheless, I argue that our discipline has the tools to mitigate and overcome the problems that limited replicability poses.

In light of the problems brought up by the replicability crisis in scientific research, Munafò and Davey Smith argued that instead of mere replication, research should strive toward triangulation: “*This is the strategic use of multiple approaches to address one question. While each approach has its own explicit assumptions, strengths and weaknesses, results that agree across different methodologies are less likely to be artefacts*” (2018, 400). In archaeology, such an approach has been – or should have been – always part of the process to some extent.

Wylie described how confidence in archaeological interpretations

can be increased through the integration of several independently constructed lines of evidence, which would not only act as mutually reinforcing, but also as mutually scrutinizing (Wylie 1989, 2000, 2002). For example, archaeological inquiry often resorts to triangulating data from different dating methods. Likewise, the identification of specific tool-mediated human activities is more reliable when both wear analysis and residue analysis on a tool point in the same direction (e.g. Cristiani and Zupancich 2021, Li et al. 2020). Within the restricted field of research on warfare and violence one could postulate that a more refined view of the movements and the tactics involved in violent encounters could be reached by triangulating information coming from experiments and wear analysis on defensive implements (Mödlinger 2018, Molloy 2009) and offensive weaponry (Gentile and van Gijn 2019, Hermann et al. 2020b), the study of bone trauma (Brinker et al. 2018, Downing and Fibiger 2017), and skeletal robusticity patterns (Gentile et al. 2018).

At a smaller scale, knowledge refinement and inter-scrutiny is also possible within the same field of experimental (archaeology) research. Although aiming at answering related research questions, experimental setups with considerably different degrees of control and variables involved, could be considered, de facto, different methods of research one could cross-compare and scrutinize results from. By 'tacking back and forth' (sensu Wylie 1989) along the spectrum of actualism and control, cause and effect are better distinguished from correlation, confidence over the results is increased, and potential for generalisation is expanded.

For what concerns combat experiments with copper-alloy weaponry, for example, distinct setups with rather varying degrees of control produced remarkably similar traces, at least qualitatively (cf. pictures of traces in Gentile and van Gijn 2019, Hermann et al. 2020a, Knight 2019, O'Flaherty et al. 2011). Despite the limited amount of repetitions within each set-up that afflicts our field (see Dolfini and Crellin 2016), such overall convergence increases the confidence in the interpretation of the results. At this moment, although much is left to explore when it comes to specific trace formation dynamics, there is wide consensus that – for instance – certain notches and dents found on archaeological metal weapons are strong indicators of use in combat. Triangulating different set-ups might also be seen as a way to achieve more generalization thanks to the introduction of minor variability, such as slightly different combat movements, or differences in the manufacture of the replicas. Besides triangulation with other experimental setups, multi-stage experiments, like the one with bronze spears described above, represent an attempt to increase confidence in the interpretations by incorporating a small degree of triangulation (and a level of cross-examination) already within the same experimental layout (see also Hermann et al. 2020b for a similar attempt).

Last, obviously, assessment of the results should also pass by the observation of the archaeological record, tacking back and forth between the ever more refined experimental reconstruction and the traces observed on the artifacts (Wylie 1989, 2002). Within the field of combat experiments, a fitting example of this approach is the constant (re) assessment of the viability of historical combat techniques as a framework for the reconstruction of prehistoric combat, through the comparison between the traces produced experimentally and those found on archaeological weaponry (cf. Gentile and van Gijn 2019, and Hermann et al. 2020b).

Nevertheless, while convergence of results is encouraging and should certainly be valued, it is paramount to direct our efforts also towards tackling ambiguity. Even when distinct lines of investigations point in the same direction, the strength of an interpretative hypothesis is assessed not only by the amount of supportive empirical evidence but also by its resilience to continuous scrutiny and attempts to falsify it (Popper 2002). In wear-traces experiments on metalwork implements and weaponry, even when replication is achieved and a solid consensus is reached on how variables X and Y influence the development of trace Z, currently we still know too little about whether other phenomena or

activities could also have produced Z. Although replication of the same experiments is an essential part of a collective research process, this effort should not be pursued at the expenses of the exploration of counterfactual inference: establishing whether different sets of variables would not produce the same result.

Tackling ambiguity could, for example, strengthen confidence in the study of archaeological bronze weaponry. A large portion of the Bronze Age swords and spears currently stored in museum collections are single finds and hoards discovered fortuitously during farming works, the dredging of rivers, or found by detectorists and enthusiasts (cf. Fontijn 2002, Verlaeck 1996, York 2002). Furthermore, it is not rare for these items to stay in private collections for a long time before being available for study and often undergo invasive 'restoration' processes. While wear analysis can provide a solid contribution to the reconstruction of the biography and the use of the weapons prior to their deposition, the lack of knowledge on how to decode the post-depositional life of these items can hinder or reduce the potential of the analysis (cf. Amkreutz et al. 2019). It follows that, the effects that more recent agents, such as – among others – mechanical ploughs or dredging machines, have on the preservation of archaeological wear traces and on the formation of additional marks needs to be addressed urgently. Reducing the level of ambiguity becomes even more relevant considering the custom of depositing weaponry in a bent and fragmented state during the Bronze Age and Iron Age and the necessity to distinguish such culturally meaningful acts from recent modern damage (Knight 2019, Mörtz 2018, Nebelsick, 2000). Directly related to this issue is the much-understudied relationship between wear traces and corrosion and how the latter can mask, alter, or even mimic wear traces (Horn and von Holstein, 2017). A better understanding of the effect of all the post-depositional processes (including curation history) have on wear traces is paramount to reduce ambiguity, broaden the knowledge of the life-path of the items, and strengthen archaeological interpretations.

5. Conclusion

This paper discussed challenges and workarounds in the design of experiments attempting to conciliate replicability and reproducibility with a sufficient amount of realism. Attention was drawn to the issues surrounding the selection of contact materials and their effect on the replicability of the experiments, as well as their value for generalization. Finally, the challenges that experimental layouts leaning towards realisms pose in terms of ethics, safety, and sustainability were examined.

In a relatively young field such as experimental archaeology of combat, detailed documentation and the consequent replication of experimental setups is of great importance. At the same time, while perfect replication remains challenging, I have argued that triangulation of different experimental programs with similar goals contributes to compensate issues such as a generally low number of repetitions per experiment and increases the potential towards generalization of the results. For these reasons, and thanks to the increasing number and quality of experiments carried out, as well as to the active efforts towards setting up common frameworks and guidelines (Dolfini and Crellin 2018), the development of the discipline appears promising. Nevertheless, it is paramount that we also direct efforts towards tackling ambiguity. Strengthening confidence in the interpretations of wear traces cannot only rely on the replication of experimental layouts but needs to be grounded also on the investigation of possible counterfactual evidence.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

I thank the editors of this special issue for giving me the opportunity to present and discuss this paper. Furthermore, I would like to thank the two anonymous reviewers whose comments greatly improved this paper. I am indebted to Casper van Dijk, Jaap Hogendoorn, and Oskar Ter Mors for their expert advice and for making the experiments possible. I am grateful to Annelou van Gijn for her comments and advice. Similarly, I thank Bastiaan Steffens and Anna Friedler for their comments on the earliest drafts of this manuscript and Sasja van der Vaart-Verschouf for language proofing the final text. All opinions and errors rest with the author.

Funding

This publication is part of the project 'Weapons of choice – Exploring attitudes towards mortality in Early Europe' (project number 322-60-010) of the research programme 'Phds in the Humanities' which is (partly) financed by the Dutch Research Council (NWO).

References

- Amkreutz, L., Fontijn, D., Gentile, V., 2019. Down by the River. A Late Bronze Age full-hilted sword from the river Meuse near Thorn and Wessem. In: Arnoldussen, S., Ball, E.A.G., van Dijk, J., Norde, E., de Vries, N. (Eds.), *Metaaltijden 6. Bijdragen in De Studie Van De Metaaltijden*. Sidestone Press, Leiden, pp. 113–122.
- Anderson, K., 2011. Slashing and thrusting with Late Bronze Age spears: Analysis and experiment. *Antiquity* 85 (328), 599–612. <https://doi.org/10.1017/S0003598X00067983>.
- Baker, M., 2016. 1,500 scientists lift the lid on reproducibility. *Nature* 533 (7604), 452–454. <https://doi.org/10.1038/533452a>.
- Bridgford, S., 2000. *Weapons, warfare and society in Britain 1250–750 BC*. University of Sheffield. Unpublished PhD thesis.
- Brinker, U., Harten-Buga, H., Staupe, A., Jantzen, D., Orschiedt, J., 2018. Perimortem Lesions on Human Bones from the Bronze Age Battlefield in the Tollense Valley: An Interdisciplinary Approach. In: Dolfini, A., Crellin, R.J., Horn, C., Uckelmann, M. (Eds.), *Prehistoric Warfare and Violence: Quantitative and Qualitative Approaches*. Springer, Cham, pp. 39–60. https://doi.org/10.1007/978-3-319-78828-9_3.
- Callahan, E., 1999. What is Experimental Archaeology? In: Westcott, D. (Ed.), *Primitive Technology: A Book of Earth Skills*. Gibbs Smith, pp. 4–6.
- Crellin, R.J., Dolfini, A., Uckelmann, M., Hermann, R., 2018. An experimental approach to prehistoric violence and warfare. In: Dolfini, A., Crellin, R.J., Horn, C., Uckelmann, M. (Eds.), *Prehistoric Warfare and Violence: Quantitative and Qualitative Approaches*. Springer, Cham, pp. 279–305. https://doi.org/10.1007/978-3-319-78828-9_9.
- Cristiani, E., Zupancich, A., 2021. Sandstone Ground Stone Technology: a Multi-level Use Wear and Residue Approach to Investigate the Function of Pounding and Grinding Tools. *Journal of Archaeological Method and Theory* 28 (2), 704–735. <https://doi.org/10.1007/s10816-020-09488-1>.
- Dolfini, A., Collins, R., 2018. Modelling Physical and Digital Replication: Bridging the Gap between Experimentation and Experience. *Open Archaeology* 4 (1), 36–49. <https://doi.org/10.1515/opar-2018-0002>.
- Dolfini, A., Crellin, R.J., 2016. Metalwork wear analysis: The loss of innocence. *Journal of Archaeological Science* 66, 78–87. <https://doi.org/10.1016/j.jas.2015.12.005>.
- Downing, M., Fibiger, L., 2017. An experimental investigation of sharp force skeletal trauma with replica Bronze Age weapons. *Journal of Archaeological Science: Reports* 11, 546–554. <https://doi.org/10.1016/j.jasrep.2016.12.034>.
- Fontijn, D., 2002. *Sacrificial landscapes. Cultural Biographies of person, objects and "natural" places in the Bronze Age of the southern Netherlands*. Leiden University Faculty of Archaeology, Leiden.
- Gentile V., van Dijk, C. & Ter Mors, O. (in prep). *Multi-stage Experiments in Bronze Age spear combat: insights on combat contexts and wear formation*.
- Gentile, V., van Gijn, A., 2019. Anatomy of a notch. An in-depth experimental investigation and interpretation of combat traces on Bronze Age swords. *Journal of Archaeological Science* 105, 130–143. <https://doi.org/10.1016/j.jas.2019.02.004>.
- Gentile, V., Sparacello, V.S., D'Ercole, V., Coppa, A., 2018. Martial Practices and Warrior Burials: Humeral Asymmetry and Grave Goods in Iron Age Male Inhumations from Central Italy. In: Dolfini, A., Crellin, R.J., Horn, C., Uckelmann, M. (Eds.), *Prehistoric Warfare and Violence: Quantitative and Qualitative Approaches*. Springer, Cham, pp. 61–83. https://doi.org/10.1007/978-3-319-78828-9_4.
- Hermann, R., Crellin, R.J., Uckelmann, M., Wang, Q., Dolfini, A., 2020a. Bronze Age Combat: an experimental approach. BAR Publishing, Oxford.
- Hermann, R., Dolfini, A., Crellin, R.J., Wang, Q., Uckelmann, M., 2020b. Bronze Age Swordsmanship: New Insights from Experiments and Wear Analysis. *Journal of Archaeological Method and Theory* 27 (4), 1040–1083. <https://doi.org/10.1007/s10816-020-09451-0>.
- Horn, C., von Holstein, I.C.C., 2017. Dents in our confidence: The interaction of damage and material properties in interpreting use-wear on copper-alloy weaponry. *Journal of Archaeological Science* 81, 90–100. <https://doi.org/10.1016/j.jas.2017.04.002>.
- Jaquet, D., Sørensen, C.F., Cognot, F., 2015. Historical European Martial Art a crossroad between academic research, martial heritage re-creation and martial sport practices. *Acta Periodica Duellatorum* 3 (1), 5–35. <https://doi.org/10.1515/apd-2015-0001>.
- Knight, M.G., 2019. Going to Pieces: Investigating the Deliberate Destruction of Late Bronze Age Swords and Spearheads. *Proc. Prehist. Soc.* 85, 251–272.
- Kuijpers, M., 2017. *An Archaeology of Skill: Metalworking Skill and Material Specialization in Early Bronze Age Central Europe*, first ed. Routledge, London. <https://doi.org/10.4324/9781315196022>.
- Lammers-Keijsers, Y.M.J., 2005. Scientific Experiments: A Possibility? Presenting a General Cyclical Script for Experiments in Archaeology. *EuroREA* 2, 18–24.
- Lee, Y., Hwang, K., 2002. Skin thickness of Korean adults. *Surgical and Radiologic Anatomy* 24 (3–4), 183–189. <https://doi.org/10.1007/s00276-002-0034-5>.
- Li, W., Tsoraki, C., Yang, Y., Xin, Y., van Gijn, A., 2020. Plant foods and different uses of grinding tools at the Neolithic site of Tanghu in Central China. *Lithic Technology* 45 (3), 154–164. <https://doi.org/10.1080/01977261.2020.1755789>.
- Mathieu, J.R., 2002. Experimental Archaeology: Replicating past objects, behaviors, and processes. In: Mathieu, J.R. (Ed.), *Experimental Archaeology: Replicating Past Objects, Behaviors, and Processes*. Archaeopress, Oxford, pp. 1–11. <https://doi.org/10.30861/9781841714158>.
- Mödlinger, M., 2018. Body Armour in the European Bronze Age. In: Dolfini, A., Crellin, R.J., Horn, C., Uckelmann, M. (Eds.), *Prehistoric Warfare and Violence: Quantitative and Qualitative Approaches*. Springer, Cham, pp. 177–198. https://doi.org/10.1007/978-3-319-78828-9_9.
- Molloy, B., 2008. Martial arts and materiality: a combat archaeology perspective on Aegean swords of the fifteenth and fourteenth centuries BC. *World Archaeology* 40 (1), 116–134. <https://doi.org/10.1080/00438240701843611>.
- Molloy, B., 2009. For Gods or men? A reappraisal of the function of European Bronze Age shields. *Antiquity* 83, 1052–1064. <https://doi.org/10.1017/S0003598X00099348>.
- Molloy, B., 2010. Swords and swordsmanship in the Aegean Bronze Age. *American Journal of Archaeology* 114 (3), 403–428. <http://www.jstor.org/stable/25684288>.
- Mörtz, T., 2018. Violence and ritual in late bronze age Britain: weapon depositions and their interpretation. In: Horn, C., Kristiansen, K. (Eds.), *Warfare in Bronze Age Society*. Cambridge University Press, pp. 168–188.
- Munafò, M.R., Davey Smith, G., 2018. Robust research needs many lines of evidence. *Nature* 553 (7689), 399–401. <https://doi.org/10.1038/d41586-018-01023-3>.
- Nebelsick, L. D. (2000). Rent Asunder, Ritual Violence in Late Bronze Age Hoards. In C. F. E. Pare (Ed.), *Metals Make the World go Round, The Supply and Circulation of Metals in Bronze Age Europe. Proceedings of a conference held at the University of Birmingham in June 1997*. Oxford: Oxbow Books.
- O'Flaherty, R., 2007. A weapon of choice – experiments with a replica Irish Early Bronze Age halberd. *Antiquity* 81 (312), 423–434. <https://doi.org/10.1017/S0003598X00095284>.
- O'Flaherty, R., Gilchrist, M.D., Cowie, T., 2011. Ceremonial or deadly serious? New insight into the function of Irish Early Bronze Age halberds. In: Uckelmann, M., Modlinger, M. (Eds.), *Bronze Age Warfare: Manufacture and Use of Weaponry*. Archaeopress, Oxford, pp. 39–52.
- Outram, A.K., 2008. Introduction to experimental archaeology. *World Archaeology* 40 (1), 1–6. <https://doi.org/10.1080/00438240801889456>.
- Popper, K., 2002. *The Logic of Scientific Discovery*, Second ed. Routledge, London. <https://doi.org/10.4324/9780203994627>.
- Schenck, T., 2011. Experimenting with the Unknown. In: Petersson, B., Narmo, L.E. (Eds.), *Experimental Archaeology, Between Enlightenment and Experience. Department of Archaeology and Ancient History Lund University, Lund*, pp. 87–98.
- Sokolov, V.E., Danilkin, A.A., 1979. A skin shield in Siberian roe males (*Capreolus capreolus pygargus* Pall.). *Mammalia* 43 (3). <https://doi.org/10.1515/mamm.1979.43.3.391>.
- Soriano Llopis, I., & Gutierrez Sáez, C. (2009). Use wear analysis on metal: the influence of raw material and metallurgy processes. *2nd International Conference. Archaeometallurgy in Europe.*, 115–124.
- Uckelmann, M., 2012. *Die Schilde der Bronzezeit in Nord-, West- und Zentraleuropa*, Stuttgart, Steiner.
- Van Dijk, C. 2020. *Feeling a Bronze Spear. A Phenomenological approach for understanding the combative use of Bronze Age spears (1800-600 BC) during formalised fights*. Leiden: MA thesis, Faculty of Archaeology University of Leiden.
- Verlaeck, K., 1996. *Between river and barrow. A reappraisal of Bronze Age metalwork found in the province of East-Flanders*. Archaeopress, Oxford.
- Wylie, A., 1989. Archaeological Cables and Tacking: The Implications of Practice for Bernstein's 'Options Beyond Objectivism and Relativism'. *Philosophy of the Social Sciences* 19 (1), 1–18. <https://doi.org/10.1177/004839318901900101>.
- Wylie, A., 2000. Questions of Evidence, Legitimacy, and the (Dis)Unity of Science. *American Antiquity* 65 (2), 227–237. <https://doi.org/10.2307/2694057>.
- Wylie, A., 2002. *Thinking from Things. Essays in the Philosophy of Archaeology*. University of California Press, Berkeley.
- York, J., 2002. *The Life Cycle of Bronze Age Metalwork From the Thames*. Oxford Journal of Archaeology 21 (1), 77–92. <https://doi.org/10.1111/1468-0092.00150>.