Journal of Archaeological Science: Reports 53 (2024) 104411

Contents lists available at ScienceDirect

Journal of Archaeological Science: Reports

journal homepage: www.elsevier.com/locate/jasrep

Controls, conceits, and aiming for robust inferences in experimental archaeology

Metin I. Eren [a,b], David J. Meltzer [c]

[a] Department of Anthropology, Kent State University, Kent, OH 44242, USA
[b] Department of Archaeology, Cleveland Museum of Natural History, Cleveland, OH 44106, USA
[c] Department of Anthropology, Southern Methodist University, Dallas, TX 75205, USA

ABSTRACT

Although experimental archaeology in some form has existed for more than a century, in the last couple of decades it has matured as a useful approach for making inferences and developing testable hypotheses about the archaeological record and the evolution of technology. However, despite several theoretical and methodological advances and growing consensus about best practices, problematic issues persist. One problem with which archaeologists must still contend – and our focus here – is understanding the inferential limits of an archaeological experiment. Using historic and modern experimental disagreements regarding North American beveled points as a case study, we explore issues of experimental controls and their tradeoff with experimental realism. We conclude with a discussion of several suggested ways archaeologists can temper inferences made from experiments.

1. Introduction

Experimental studies have a centuries-long history in archaeology, and in the last several decades have become increasingly more frequent, increasingly more sophisticated, and increasingly applied in archaeological inference (e.g., Bebber et al., 2023; Calandra et al., 2020; Coles, 1973,1979; Coppe et al., 2019; Eren et al., 2016a; Eren and Bebber 2020; Ferguson, 2010 and references therein; Iovita et al., 2014; Jennings et al., 2021; Key, 2016; Key and Lycett 2017a; Li et al., 2023; Lin et al., 2018; Lycett and Chauhan, 2010; Magnani et al., 2014, 2019a, 2019b; Martellotta et al. 2023; Mesoudi and O'Brien 2008; Milks et al., 2016, 2019; Muller et al., 2022; Neill et al., 2022; Outram, 2008; Pargeter et al., 2023a, 2023b; Rezek et al., 2016; Schillinger et al., 2014,2015; Schiffer, 2013; Sisk and Shea, 2009; Stemp, 2016; Wilkins et al., 2012). Nonetheless, certain problematic elements – and attitudes – have also persisted over time that limit the effectiveness of many experimental efforts. Ironically, that includes a lack of understanding of just what the inferential limits are to an archaeological experiment, and how much confidence in the conclusions one can draw from them given the customary parameters of most experiments. We explore these issues here, focusing on issues of experimental controls and their tradeoff with experimental realism, and experimental/experimenter conceit (the latter referring to an experimenter declaring an experiment conclusive on their own authority, such as their skill as flintknappers or hunters). To put it another way, in a world of archaeological equifinality, where an unknown number of processes might have resulted in the same product, can we really be that certain our experimental inferences are correct? And, if not, how do we better temper our results and inferential reach?

To illustrate these conceptual issues, we focus on a series of experiments conducted since the late 19th century that have sought to understand the function of a distinctive attribute of projectile points: beveling of the point blade, such that it has a rhomboid shape in cross section (Fig. 1). In North America, beveled projectile points appear in Late Paleoindian times, notably on Dalton period projectile points (and its regional variants) as well as on some Early Archaic projectile point forms (Lipo et al., 2012; Pettigrew et al., 2015). Beveling of projectile points largely disappears thereafter, save for a brief re-emergence in Early Woodland times (Lipo et al., 2012).[1]

The function and/or aerodynamic consequences of beveling has eluded archaeologists, or rather has eluded agreement among archaeologists. The principal question, which has been the subject of multiple experimental studies, is whether beveling would cause a projectile point to rotate in flight. Some claim to have conclusively shown experimentally that is the case, while others' experiments are claimed to decisively reject that proposed effect. The lack of experimental agreement might be a consequence of the nature of the different experimental approaches taken or, perhaps, the fact that under certain experimental conditions

E-mail addresses: meren@kent.edu (M.I. Eren), dmeltzer@smu.edu (D.J. Meltzer).

[1] There are also in North America beveled bifaces, most notably the so-called Harahey knives found in Late Prehistoric sites on the Great Plains, and thought to have been bison butchering tools.

https://doi.org/10.1016/j.jasrep.2024.104411
Received 26 October 2023; Received in revised form 15 January 2024; Accepted 21 January 2024

Fig. 1. An example of a beveled notched projectile point (knapped by M.I.E.).

beveling sometimes makes points spin, but not under other experimental conditions (Key and Lycett 2017b). It should also be acknowledged that the spinning (or not) of a point in flight could be a random or incidental occurrence with respect to its being beveled. That is, the beveling of a projectile point was not functional in any ballistic sense (intended or not), but was instead intended to cause larger entry wounds in prey (Ashby, 2007; Pettigrew et al., 2015). Or perhaps instead of it being a functional attribute, it was a technological by-product of resharpening of the point (Bradley, 1997; Goodyear, 1974; 1997; Pettigrew et al., 2015; Sollberger, 1971), or the result of an overshot flake mistake, or even the consequence of expediently knapping a bifacial point on a flake that already possesses a twisted or beveled morphology (e.g. Eren et al., 2016b, 2021a). It may well be, of course, that the beveling of points had both functional and technological elements (e.g. Smallwood et al., 2020). Archaeological equifinality, again.

These alternative functional and non-functional proposed explanations for beveled points illustrate the frustrating fact that even if one specific function or aerodynamic action is experimentally supported, subsequently inferring whether that was intentionally selected, or an incidental byproduct of another selected function, production strategy, or even drift, can be difficult if not impossible to determine (Eren et al., 2022a). In other words, given the resolution of the archaeological record (Conrad et al., 2023; Perrault 2020), pinpointing a technology's precise mechanism or moment of innovation, or the reason for its origin, is challenging. We will likely never know whether past peoples intentionally designed a new, useful technology (or technological feature) or instead reaped the benefits of an unintentional technological accident or discovery. Past peoples need not have even been fully conscious of the benefits of their technology or understood the reasons why their technology worked (e.g., Harris et al., 2021) – and whether they did or not is also likely forever beyond archaeological purview. It is important to keep these latter points in mind, even though our focus here is on the broader question of how or whether experimental approaches can establish that an artifact possesses a particular function or conveys a selective advantage, regardless of whether it was intentional or incidental or even recognized at the time.

2. Experiments on the function of projectile point beveling

Thomas Wilson, Curator of the Division of Prehistoric Archeology in the United States National Museum (Smithsonian Institution) in the last decades of the 19th century (Meltzer, 2015), conducted the first set of experiments seeking to understand beveling, in one of the earliest archaeological experiments on record. It involved testing his suspicion beveling would cause projectile points to rotate in flight (Wilson 1898, 1899). He may have had the idea that this "rifling" would increase the accuracy of the armature in flight, although he does not say so.

Selecting more than a dozen beveled artifacts from the U.S. National Museum collections, Wilson attached each to an arrow shaft and went to the top of the tower of the Smithsonian building, where he either let them fall straight down,[2] or he launched them into the air in different directions. In addition, he fashioned a means by which the points could be sent through water in a large tub, placed in front of a pipe of rapidly flowing air, or in front of a high-speed fan. In all experiments, which were evidently done repeatedly, the result was the same: he "found a universal rotation" of the points through the air (or water) (Wilson 1899:932). He suspected the rotation was intended by the maker of the points, though he was puzzled why an 'arrow maker' would do so, given that twisting the feathers on the shaft would accomplish the same purpose with less labor (he obviously thought these were arrow points as opposed to dart points).

Decades later, reading a report that described an arrow point as having "definite rifling or beveling which would presumably ensure whirling in flight" (Campbell and Ellis, 1952: 217), Arthur George Smith could scarcely conceal his disdain: "this is an amazing statement to find in a serious report made by two competent modern archaeologists" (Smith, 1953:169). Smith had long doubted that "fable about 'rotary points'" (he described it as "old when I was a lad"), since there was not enough force generated against the surface of the bevels to spin the shaft of an arrow – let alone spin larger and heavier dart points. He reported on his own series of experiments which involved shooting arrows (with non-feathered shafts to eliminate featherings' rotational effects) with beveled points into the air, as well as having one of his friends shooting arrows at or past him (which he could dodge), observing them to determine whether the points rotated in flight. He saw no such evidence and concluded that "the beveling had no effect on the flight of the arrow" (Smith, 1953:269), and more likely was due to the specimens having been resharpened while hafted.

Two sets of experiments came to two opposite conclusions. Yet, each had overwhelming confidence in his results' archaeological applicability. Wilson (1899:933) claimed:

> …these experiments were pushed to such extent and in such number, with such repetition of the same result, as to be conclusive that, whatever may have been the intention of the maker of the arrow-points, the fact was that in their flight through the air the beveled edges produced the rotary motion.

Smith (1953:270) was just as certain:

> The fable that beveled points were made in that manner to spin an arrow in flight is in the same category as the fable about the Mound Builders tempered copper, and the one about chipping arrowheads from red-hot flint with an icicle. Let us forget them or leave them to the writers of filler pieces for the Sunday Supplement.

With a century of hindsight (and acknowledging the unfairness of presentism), it is easy to see that their unbounded confidence was not warranted. We can credit both with recognizing the importance of varying their experimental conditions to determine if the results could

[2] The tower was part of the Smithsonian's 'Castle' built in 1855. The specimens Wilson dropped from the Castle tower were "carried only with their own gravity," and thus would have accelerated at the rate of 9.8 m/s^2. As a ballpark estimate, the tower is 44 m tall at the top, though presumably Wilson was not standing atop the tower, but below on one of the viewing platforms. Assuming he was at an elevation of, say, 40 m above ground level, the terminal velocity of the dropped points would have been 28 m/s, which is within the range of the velocity of a thrown dart.

be replicated (Wilson did not depend solely on dropping, launching, or putting the points in water tanks or wind tunnels: he tried all of them; Smith sought to observe rotation from different angles).[3]

Yet, other elements of these early experiments are problematic. Their experimental parameters, for example, were largely uncontrolled: neither provides specifics regarding sample size, either in terms of numbers of specimens or the number of trials. Neither describes how, or at what velocity, the arrows were launched (or, for example, the air speeds of the fans in Wilson's experiments). Wilson does not indicate the beveled points' raw material (though it appears he used actual specimens from the museum collections); Smith used plate glass for his specimens. Neither Wilson nor Smith provided the arrow or dart shafts' wood type, or the forms or mass of any of these items. Non-beveled points, which might have provided experimental comparisons and controls, were not included in either experiment. Finally, absent any actual measuring devices, and because they had to make their assessments visually as the projectile points fell or flew away (Wilson) or came toward them through the air (Smith[4]), it is unlikely their observations of whether and how rapidly the points rotated are accurate.

Might we today, with the more sophisticated experimental approaches now in use, be more confident in our experimental results, and apply those to explain patterns of artifacts or features of the archaeological record? Do experimental demonstrations mean that archaeologists have definitively replicated past behaviors? To consider these questions, we turn to two modern experiments that also focused on whether beveled points cause projectile point rotation, ones by Lipo et al. (2012), and Pettigrew and colleagues (Pettigrew et al., 2015). Each approached the question in very different ways.

Before conducting any experiments with actual specimens, Lipo et al. (2012) first simulated the beveled projectile point rotation performance using Computational Fluid Dynamics (CFD) (a computer simulation can be considered a type of highly controlled experiment). As they note, CFD is a means of studying the properties of fluids like gases or liquids using numerical methods to approximate the complex conditions involved in the flow of fluids around solid objects. Subjecting digital representations of beveled points to theoretical wind speeds spanning 5 to 60 m/s (for comparison and broadly speaking, the velocity of a thrown dart might range from 20 to 25 m/s, that of an arrow 40–50 m/s), their CFD analysis showed that as wind speed increases, so too the forces on each beveled face of the point, potentially inducing the point to spin. To examine empirically the question of beveled-point rotation, Lipo et al. (2012) then ran a series of experiments using modeled (acrylic) and prehistoric beveled bifaces in a low-speed wind tunnel with a maximum wind speed of 30 m/s (Lipo et al., 2012: 779). Their empirical results affirmed their simulated findings: both replica and real beveled bifaces rotated in the wind tunnel, and this rotation increased as wind speed increased.

Pettigrew et al. (2015) reject the premise of Lipo et al's approach, arguing that "the problem with all controlled experiments and theoretical models is that they may be too far from the practical reality of a prehistoric technology; they necessarily examine too few of a suite of interacting variables" (Pettigrew et al., 2015:591). Accordingly, they also reject the conclusions Lipo et al. drew from those experiments.

For their part, Pettigrew et al. conducted a series of so-called 'naturalistic' experiments "more relevant to prehistoric dart use" (Pettigrew et al., 2015:594) using a CASIO EX-F1 high speed camera to assess whether human-thrown atlatls tipped with beveled and unbeveled darts rotated. The points did not rotate, at least not in a consistent manner: "it is clear that different darts – with a conical point or no point, with an unbeveled point, and with a beveled point – all behave unpredictably: some spin consistently in one direction, some in the other, and some reverse their direction of spin flight" (Pettigrew et al., 2015:595). Any spinning that occurred they deemed more likely to be a function of the oscillation of the dart, and not because of the beveling of the point or the fletching of the dart (Pettigrew et al., 2015:597). They raise the possibility – and tested it with melons and a hog carcass – that beveling may have caused points to rotate on impact, but otherwise like Smith (1953) considered resharpening the best explanation for beveling (Pettigrew et al., 2015:599).

Thus, on the matter of rotation in flight, and just as Wilson (1898) and Smith (1953) had, Lipo et al. (2012) and Pettigrew et al. (2015) arrived at opposite and seemingly contradictory conclusions, and attribute beveling to different purposes. Similarly, like Wilson (1898) and Smith (1953), Lipo et al. (2012) and Pettigrew et al. (2015) are each fully convinced of the validity of their results and their findings' broad archaeological applicability. According to Lipo et al., (2012:787):

> We have shown that beveling causes pointed bifaces to spin in flight. This has been demonstrated both theoretically and by wind-tunnel experimentation. In-flight rotation offers benefits in the form of increased accuracy for ballistics shafts that have a mass range consistent with thrown spears and atlatl-launched darts sufficient to explain the fixation of the trait. […] In the cultural context of the midwestern and southeastern United States, beveling appears to be an early, if not the first, adaptation to transform the long-handled knife/stabbing tool represented by the Clovis point and its kin into an efficient casting instrument.

Pettigrew et al., having dismissed Lipo et al.'s (2012) wind tunnel experiment as "*completely irrelevant to the behavior of real projectiles*," argue that "*the archaeological conclusions derived from it must therefore be discarded*" (Pettigrew et al., 2015: 593, emphasis in the original). Beveled points do not spin a projectile in flight, based on their naturalistic experiments. Lest there be questions as to whether those experiments could have been flawed by poorly made or used equipment, they cite their "several years of experience" making and using such atlatls, along with their observations of other atlatl throwers they have seen live or in videos (Pettigrew et al., 2015: 595).

So would beveled points have spun in the air when launched by past peoples? We cannot say. What we can say is that the Lipo-Pettigrew disagreement highlights two important challenges archaeological experimenters face: namely, the matter of experimental controls, and the danger of experimental conceit. Taken together, these challenges suggest that one of the biggest mistakes an experimental archaeologist can make is to be too confident in the inferential power of any one experiment, or perhaps even a series of experiments.

3. Thinking about experimental controls, conceits, and generalizability

Lipo et al.'s (2012) investigation of beveled points and in-flight rotation involved computer simulations and experiments in a wind tunnel: these are highly controlled experiments. Much has been written about experimental control in archaeology (e.g. Eren et al., 2016a; Lin et al., 2018), but briefly it can be understood as an element of a research design strategy. "More control" means minimizing and/or constraining as much as possible the effects of extraneous variables in an experiment, in order to more precisely observe and measure the causal relationship between the variables of interest. "Less control" means permitting more experimental variable interaction, which may provide more realism, *but only in the sense that the variables are free to interact in a real and complex world – not that one is recreating a past reality*. However, this latter point can be lost if "less controlled" archaeological experiments are referred to as 'naturalistic' or 'actualistic,' as they often are (including in the past by

[3] One could perhaps also make the argument that Wilson recognized the importance of conducting experiments that are both less controlled, which provide increased variable interaction, as well as experiments that are more controlled, which isolates specific variables (Calandra et al., 2020; Eren et al., 2016a; Lin et al., 2018; Mesoudi, 2011).

[4] Smith painted the bevels of his glass points black, to enhance his chances of seeing the flash of the glass faces of the points if indeed they rotated (Smith, 1953:269).

the senior author). Such terms are misnomers, in so far as no experiment should be interpreted as representing an actual event in the past (a meaning that Binford never linked to what he termed actualistic studies [e.g. Binford, 1981:27]). Further, and as Conrad et al. (2023) recently discussed, there is little that is natural about 'naturalistic' experiments – they can be just as contrived as more controlled experiments, with as many if not more unknowns. Rather than use euphemisms like 'naturalistic' or 'actualistic,' we suggest experiments would be better referred to simply as 'more' or 'less' controlled, relative terms that convey more useful information about the number of variables and interactions in the experimental equation.

A key advantage of carefully applying more experimental controls is that one can ascertain, sometimes even with a limited number of test trials, causal relationship(s) among observed test variables (Lin et al., 2018). Yet, highly controlled experiments are not without drawbacks. Lin et al., (2018:676) point out that "experimental setups often require considerable time and financial investment in the construction of mechanical apparatus and the preparation of identical test samples."[5] Highly controlled experiments also may remove important interactions between other variables that possibly were operable in the past, and thus lead to erroneous answers to the archaeological questions being asked (Eren et al., 2016a; Lycett and Eren, 2013; Pettigrew et al., 2015). Finally, achieving more control in an experiment may introduce variables or conditions that were not operable in the past, lessening the direct relevance of an otherwise well controlled experiment.

These can be taken to farcical extremes, as when a team of experimenters, understandably reluctant to stampede a bison herd over a jump, instead rolled car and truck tires down a hill to simulate how long and far fast-moving bison may have been airborne when stampeded, and where their "first bounce" would have occurred (https://news.arizona.edu/story/new-technologies-and-tires-reconstruct-ancient-bison-hunts; https://www.archaeology.org/issues/155–1411/letter-from/2587-letter-from-montana-buffalo-jumps). While tires may be easier to record and control than bison, tires are so far removed from archaeological reality their archaeological relevance is difficult to see: bison, for example, do not bounce or roll like a tire.

Given these drawbacks, is the controlled experiment of Lipo et al. (2012) "completely irrelevant to the behavior of real projectiles" as Pettigrew et al. (2015) assert? Hardly. While Lipo et al (2012) do not demonstrate that beveling *causes* rotation in all possible circumstances, their controlled experiment demonstrates that the interaction of beveled points with fluids can nonetheless under certain circumstances *contribute* to projectile rotation. While that effect may in some cases be overwhelmed by other variables unaccounted for in these experiments and that might operate in the natural world, this is not grounds for asserting that "beveled points do not spin a projectile in flight" (Pettigrew et al., 2015:599). They can, it is just as yet not fully understood where and when and under what conditions that occurs.

The distinction between *causes* versus *contributes* is an important one that helps to situate the value of controlled experiments in archaeological inference. For example, a controlled experiment that demonstrates that a thicker platform contributes to larger stone flake sizes (e.g. Dibble and Rezek, 2009) should not be taken to mean that a thicker platform causes larger flakes in all instances. Instead, a controlled experimental finding such as this shows that platform thickness plays a role in flake size. When platform thickness interacts with some variables – for example, allowing a direct, straight hammer strike when the core is firmly supported – its contribution to flake size may be realized or even augmented. In other instances, as with a glancing, sideways strike when the core is loosely supported (as if knapping an *éclat dèbordant*), its contribution may be lessened, perhaps even to a degree that is not noticeable. But that does not mean the contribution of platform thickness to flake size is absent or altogether inconsequential.

We have ourselves grappled with the distinction between factors that 'cause' and factors that 'contribute' in highly controlled experiments (e.g. Mraz et al., 2019; Eren et al., 2021b,2022b), as for instance in our investigations into the relationship between Clovis point fluting and impact durability (Thomas et al., 2017; Story et al., 2019). In that experiment we compared fluted and non-fluted points via engineering simulations (static, linear finite element modeling and discrete, deteriorating spring modeling), and empirical assessment via an Instron Materials Tester (displacement-controlled axial-compression). Although our points were made from Georgetown (Edwards Formation) chert, a toolstone routinely used by Clovis and later Paleoindian groups (e.g. Bever and Meltzer, 2007), the specimens were ground rather than knapped to ensure all forms were alike in shape and size except for the presence or absence of a flute. Both sets of analyses suggested that upon impact fluted points possess a greater chance of undergoing stress redistribution, such that after a certain amount of tip damage, the stress, and thus damage, relocates to the base that was rendered thin and brittle after fluting. Relative to non-fluted points, this damage relocation increases a fluted point's overall resilience in terms of energy absorbed, the time before catastrophic breakage, and the length remaining intact until that moment of breakage (Thomas et al., 2017).

Yet, these results should not be taken to mean that Clovis flutes caused increased durability in all cases, only that fluting could have contributed to increased durability. Under some conditions that contribution may be easily expressed. Under other conditions that contribution may be diminished to the point that it appears absent. But the potential functional contribution of fluting to point resilience is in principle present.

Thus, while Lipo et al.'s (2012) highly controlled experiments are not "irrelevant," as asserted by Pettigrew et al. (2015), claims of their explanatory power and validity need to be tempered, given the many and still unknown variable interactions that under different circumstances can affect the rotary properties of beveled points. In other words, Lipo et al.'s (2012) results are not as far-reaching as they proclaim: their conclusions have outrun their experiment.

While Lipo et al.'s (2012) experiment used simulations and wind tunnels, Pettigrew et al.,'s (2015:591,600) study of beveled points and rotation employed slow motion camera footage and took advantage of their "real experience" to design their experiment. In this they tapped the "high skill levels" of the participants, their "equipment [which was] closely modeled on ethnographic and archaeological examples in situations resembling traditional usage," their "several years' experience both making and using atlatl equipment of a wide range of forms," and their skills from having "participate[d] regularly in atlatl competitions" (Pettigrew et al., 2015:595). While their accumulated experiences are surely valuable, we nonetheless cannot assume that any modern experiences, however broad they may appear, can fully or even adequately capture all the possible skills and behaviors and experiences that may have taken place over the many thousands of years that beveled dart or arrow points were in use (see discussion in Milks, 2019), let alone that they can eliminate the possibility of certain behaviors and uses *not* occurring in the past (Mullen et al., 2023).

Over 40 years ago Binford (1981) and Thomas (1986) each criticized experimental archaeologists' penchant for projecting their personal experiences, skill levels, and actions on the archaeological record: Binford (1981) called it "egographic analogy;" Thomas (1986) referred to it

[5] As an example of the scale of the financial investment that might be required, the lapidary production by https://www.Neolithics.com of 3,570 chert point possessing seven specific types of Clovis plan-view form cost over $40,000.00 and took nearly two years (e.g., Baldino et al., 2023; Eren et al., 2020, 2022c, 2023; Mika et al., 2022). Another example: the Kent State University Experimental Archaeology Laboratory's Instron Materials Tester cost over $80,000.00 and its Photron High Speed Camera cost over $35,000.00.

as “the flintknapper’s fundamental conceit.”[6] Each was aiming at the fallacy whereby an archaeologist assumes that if s/he can produce a particular result during an experiment that mimics a pattern seen archaeologically, then past people’s processes and results must have been the same. Such an assumption is problematic because all that has been shown is the possibility that a particular process could have led to a particular product (Binford, 1981:187; Thomas, 1986:621; see also Patten, 1981:12). Nor does it preclude the possibility that other processes could have produced the same archaeological pattern (back to equifinality). It is important for someone who has mastered, or thinks they have mastered, a technological skill to not become so impressed by their own abilities that they assume they are replicating past behaviors. Whether they are or are not, they cannot say, and certainly will never know. To assume otherwise is not replicating: it is imposing.

More broadly, thinking thus essentially limits our ability to see other processes in the past that are outside our present-day skills and experiences. Wobst (1978) spoke of the ‘tyranny of the ethnographic record,’ by which he meant the tendency to explain the archaeological record within the parameters of what is known from the thin slice of the human past known ethnographically. We must also guard against the ‘tyranny of the egographic record,’ by which we mean the tendency to explain the archaeological record within the parameters of our own expertise. To paraphrase Eren et al. (2016), we readily accept that in some cases hard-won expertise might provide a route to a more informed opinion than one proffered in the absence of practical experience and awareness. However, a problem arises when an experimental archaeologist, acting as his or her own informant (Shea, 2020), makes “definitive pronouncements” (Schiffer, 2016) about an experiment’s relevance and analytical reach in explaining the archaeological record. Moreover, the pronouncement is too often accepted by others (e.g. Lohse et al., 2014) based largely (if not solely) on the perceived authority or experience of the person proposing it (Binford, 1981; Eren et al., 2016a; Shea, 2020; Thomas, 1986), rather than on rigorous testing, robust data, or statistical analysis.

Pettigrew et al. (2012:599) are adamant that their experimental observations broadly “demonstrate that beveled points do not spin a projectile in flight”. Perhaps beveled points did not cause spinning under the specific circumstances of their ‘naturalistic’ experiment (more on this below). But that does not mean that beveled points would never spin in flight, simply because it has been shown by Lipo et al. (2012) that, under certain circumstances, they can. Moreover, Pettigrew et al. (2015) do not (and, indeed, cannot) show that aerodynamic circumstances analogous to the Lipo et al. (2012) experiment – absent the wind tunnel, of course – did not occur in the past. For that matter, it is irrelevant to suggest that Archaic hunters would have needed a theory of dynamics in order to consider the possibility that beveling might be advantageous to “spin and stabilize a dart in flight” (Pettigrew et al., 2015:600), any more than an outfielder in baseball needs to calculate a series of differential equations to predict the trajectory of a fly ball in order to catch it (to paraphrase Dawkins, 1976). Observant humans can detect (and take advantage) of subtle patterns even if they cannot define them mathematically (Harris et al., 2021).

And always lurking behind any archaeological inference are the unknown unknowns: for example, perhaps an atlatl, or atlatl dart design that has not preserved in the archaeological record acted in concert with beveled points to spin the whole projectile in flight. Or perhaps the bow and arrow was invented much earlier than currently thought, and lighter arrows rather than heavier darts are the correct wooden components to examine. Or perhaps past peoples had some currently undiscovered throwing actions (e.g., a special flick of the wrist) or “trick” (Eren et al., 2016a) that initially spun a projectile and point beveling simply helped to sustain that spin.

Another problem with less controlled, ‘naturalistic’ experiments, as Pettigrew et al., (2015:600) themselves acknowledge, is that they often yield results that are inconsistent from one trial to the next. In a word, the data are messy (Mullen et al., 2023). This is in some measure a function of the fewer controls on the experiment. Thus, large sample sizes are required to identify any sort of signal or pattern from experimental noise, and *very* large sample sizes are required to suggest the absence of a signal or pattern. At first glance, Pettigrew et al. (2015) appear to possess a large and statistically strong sample size in terms of their throws, with a sample size of 100. Unfortunately, their data are difficult to analyze and interpret statistically because the number of experimental conditions and components within the experiment effectively reduces that total sample size. Their 100 dart launches include five experimental settings, two types of recording equipment, three atlatl throwers, four types of atlatls, seven types of darts, and five tip types (TABLE 1). Thus, their largest comparison of beveled to non-beveled point tips in which all other conditions are held constant is just seven trials versus four trials, which is not statistically meaningful.

Statistical meaning, however, is crucial when dealing with varying evolutionary phenomena and population-scale inferences. As a thought experiment, imagine that bevel-tipped projectiles have a 60 % chance of spinning – as opposed to the 50 % chance of an unbeveled-tipped projectile spinning (we assume that both beveled and unbeveled points may spin, which is consistent with Pettigrew et al.’s experimental results in which they observed spinning, just not consistently). All it might take for a hunter to adopt a beveled point, and transmit that form to his/her children, might be a single hunting success resulting from that extra 10 % chance – he or she may not even perceive spinning, much less the increased chance of spinning. But for an experimental archaeologist to be able to recognize this increased chance amongst the noise of ‘naturalistic’ experiments, much larger effective sample sizes are necessary than Pettigrew et al. (2015) provide.

In sum, Pettigrew et al.’s (2015) experiments are potentially suggestive of how less controlled experiments can potentially obscure signals or patterns that are found in more controlled experiments. Yet, the presence of noise does not mean the absence of signal, especially when Pettigrew et al.’s (2015) small effective sample sizes preclude any strong conclusion one way or another. Even if their experimental sample sizes and results were robust and valid, their ego-graphic approach fails to acknowledge that the applicability of their study is limited to the very particular circumstances they tested, which may be entirely irrelevant to parts of the archaeological record and the human past. Moreover, experimental realism in the form of increased variable interaction should never be misinterpreted as indicative of archaeological reality. Thus, like Lipo et al. (2012), Pettigrew et al. (2015) needed to temper their inferential ambitions: their experiment cannot be taken as a definitive statement on beveled points and rotation in flight.

4. To target robust experimental inferences aim carefully

The disagreement between Lipo et al. (2012) and Pettigrew et al. (2015) – and others, for that matter (e.g. Ashby, 2007; Bradley, 1997; Eren et al., 2016b, 2021a; Goodyear, 1974; 1997; Sollberger, 1971; Smallwood et al., 2020) – regarding the function and possible in-flight aerodynamic consequences of beveling, suggests that much more experimental work is necessary. That is a good thing: replication of experiments and results is fundamental to the progress of any scientific endeavor. Not such a good thing, however, are Lipo et al.’s (2012) and Pettigrew et al.’s (2015) categorical assertions based on only a limited number of experiments, a practice in which they are hardly alone (e.g.

[6] Surprisingly, despite both Binford (1981) and Thomas (1986) having been widely cited, we could only find a handful of mentions of “egographic analogy” or “the flintknapper’s fundamental conceit” (e.g. Bonnichsen, 1983; Eve, 2018; Hester, 1993; Mange 1985; Muller, 2017; Schiffer, 2016; Shea, 2020; Stenak, 2022). Given the gravity of these logical fallacies, their disregard is concerning, and perhaps why they are still invoked today despite otherwise widespread improvements in experimental archaeology’s method, theory, and practice (Eren and Bebber, 2019; Magnani et al. 2019).

Table 1
Pettigrew et al., (2015, supplementary online materials) present data on 100 atlatl-launched darts to assess the role of point beveling in projectile rotation. Yet, within these 100 launches are 42 distinct experimental conditions. The effective sample size of each condition – whereby all reported experimental variables are held constant – is never larger than 7, and often only 1, 2, or 3. Such small samples sizes do not allow meaningful comparisons between beveled and unbeveled points. Abbreviations: OS = over-the-shoulder; BRC = Broken Roof Cave; GBI = Great Basin Inspired; SDC = Sand Dune Cave; UF = Unfletched.

Condition	Experimental variables							Sample size per condition
	Test location[1] (n = 5)	Filming method (n = 2)	Video (fps) (n = 2)	Thrower[2] (n = 3)	Atlatl type (n = 4)	Dart type	Point[3] type	
1	Cahokia 2013	OS	600	JW	SDC	Reed	Beveled stone	3
2	Cahokia 2013	OS	600	JW	SDC	Reed	None	3
3	Cahokia 2013	OS	600	JW	SDC	Reed	Unbeveled stone	3
4	Cahokia 2014	20 m	600	DP	GBI	Cane #4	Beveled stone	2
5	Cahokia 2014	20 m	600	DP	GBI	Cane #4	Beveled wood	7
6	Cahokia 2014	20 m	600	DP	GBI	Cane #4	Beveled stone	2
7	Cahokia 2014	20 m	600	DP	GBI	Cane #4	Unbeveled wood	4
8	Cahokia 2014	20 m	600	JG	GBI	Cane #4	Beveled wood	6
9	Cahokia 2014	20 m	600	JG	GBI	Cane #4	Unbeveled wood	4
10	Cahokia 2014	20 m	600	JW	Clovis[4]	Dowel	Beveled wood	2
11	Cahokia 2014	20 m	600	JW	Clovis[4]	Dowel	None	2
12	Cahokia 2014	20 m	600	JW	Clovis[4]	Dowel	Unbeveled wood	1
13	Cahokia 2014	OS	300	JG	Clovis[4]	Dowel	Beveled wood	2
14	Cahokia 2014	OS	300	JG	Clovis[4]	Dowel	None	5
15	Cahokia 2014	OS	300	JG	Clovis[4]	Dowel	Unbeveled wood	2
16	Cahokia 2014	OS	300	JG	GBI	Cane #4	Beveled stone	2
17	Cahokia 2014	OS	300	JG	GBI	Cane #4	None	3
18	Cahokia 2014	OS	300	JG	GBI	Cane #4	Unbeveled stone	2
19	Cahokia 2014	OS	300	JW	GBI	Cane #4	None	3
20	Fair 2014	OS	300	DP	GBI	UF Cane #2	Beveled wood	3
21	Fair 2014	OS	300	DP	GBI	UF Cane #2	None	1
22	Fair 2014	OS	300	DP	GBI	UF Cane #2	Unbeveled wood	3
23	Osage 2014	20 m	300	DP	GBI	UF Cane #1	Beveled wood	2
24	Osage 2014	20 m	300	JG	BRC	Willow #1	Beveled stone	1
25	Osage 2014	20 m	300	JG	BRC	Willow #1	Beveled wood	3
26	Osage 2014	20 m	300	JG	BRC	Willow #1	Unbeveled stone	2
27	Osage 2014	20 m	300	JG	BRC	Willow #1	Unbeveled wood	1
28	Osage 2014	20 m	300	JG	GBI	Cane #5	Beveled stone	2
29	Osage 2014	20 m	300	JG	GBI	Cane #5	Beveled wood	2
30	Osage 2014	20 m	300	JG	GBI	Cane #5	Unbeveled stone	2
31	Osage 2014	20 m	300	JG	GBI	UF Cane #1	Beveled stone	1
32	Osage 2014	20 m	300	JG	GBI	UF Cane #1	Beveled wood	1
33	Osage 2014	20 m	300	JG	GBI	UF Cane #1	Unbeveled stone	2
34	Osage 2014	20 m	300	JW	Clovis[4]	Cane #5	Beveled stone	1
35	Osage 2014	20 m	300	JW	Clovis[4]	Cane #5	Unbeveled stone	1
36	Osage 2014	20 m	300	JW	Clovis[4]	UF Cane #1	Beveled stone	1
37	Osage 2014	20 m	300	JW	Clovis[4]	UF Cane #1	Unbeveled stone	1
38	Survey 2014	20 m	600	DP	GBI	UF Cane #2	Beveled wood	4
39	Survey 2014	20 m	600	DP	GBI	UF Cane #2	None	2
40	Survey 2014	20 m	600	DP	GBI	UF Cane #2	Unbeveled wood	1
41	Survey 2014	OS	300	DP	GBI	UF Cane #2	Beveled wood	2
42	Survey 2014	OS	300	DP	GBI	UF Cane #2	Unbeveled wood	3

[1] The location of testing is provided, but this does not reveal other potential variables, such as temperature, wind-speed, wind-direction, etc.
[2] There is no information provided about the skill-levels of the throwers.
[3] Beyond beveling or material type, no information is provided on the point type.

[4] An example of a Clovis atlatl has not been recovered from the archaeological record.

Frison, 1989,2004; Kilby et al. 2021; Stanford and Bradley, 2012). Applicable inferences can only come from experiments after a specific result has been replicated repeatedly by independent researchers, a phenomenon has been thoroughly described and understood under several conditions, and there are robust links between the experimental and archaeological data. Even then, the inferences must be recognized for what they are: possible (and not proven) explanations, or as hypotheses worthy of testing.

Humility is also in order, recognizing the many 'experiments' our species has undertaken over the long course of prehistoric time. In that sense, then, experimental archaeologists should accept that their inferences are limited and provisional, and carefully consider their inferential ambitions. One simple and immediate solution to doing so is to carefully craft one's conclusions, and represent what they allow us to say, and what they do not allow us to say. One team's experiment is not necessarily wrong or "irrelevant" because its results come to the opposite conclusion of those of another; instead, more testing is required to understand why the two sets of results do not match. Moreover, the exploration of an experiment's limitations should become regular practice in published archaeological experiments. In some cases, this discussion might take the form of an explicit and separate "limitations section" (or sub-section). Beyond simply pointing out an experiment's drawbacks, this section could also discuss future potential experiments, alternative experimental conditions, and untested variables and variable interactions. All of this would emphasize the fact that no experiment is perfect, comprehensive, or final.

Another solution to help correct a "one and done" approach in experimental archaeology might be for archaeologists to adopt a paradigmatic classification (*sensu* Dunnell, 1971) of experimental variables approach. A paradigmatic classification is a dimensional classification procedure in which the units, i.e. classes, are defined by intersection, with each dimension (henceforth, variable) being a set of mutually exclusive alternate features (variable states). All variable states belonging to a single variable share the ability to combine with variable states of each other variable.

So, what would a paradigmatic classification of experimental variables approach look like with respect to testing whether beveled points rotated in flight? One experimental variable might be *point type*, with its three variable states being "not beveled," "one beveled edge," and "two beveled edges." A second variable could be *weapon system*, with its three variable states being "thrown spear," "atlatl and dart," and "bow and arrow." A third variable might be *fletching*, with its three variable states being "elliptical fletching," "straight fletching," and "no fletching." The intersection of these three simple variables with their three variable states results in 27 potential experiments that could be run (Fig. 2, **left**). But of course, the repetition of experiments by others is necessary; replicating each of these 27 experiments only twice each would result in a total of 81 experiments (Fig. 2, **right**). And if each of these 81 experiments only tested 10 replicated projectile specimens, that would require the knapping of 810 stone points, which would then need to be hafted onto 810 arrows shafts, and then launched. Ideally, however, each experiment would test more than just 10 specimens. The testing of 30 specimens in each of the 81 experiments would require the knapping, hafting, and shooting of 2,430 stone points.

But just how representative are these three variables (each with three variable states) of past variability? Probably not very representative. The degree of beveling; the angle of beveling; the size of points; the point raw materials; the weight, length, and wood type of spears, darts, and arrows; the type of atlatl; the type of bow; the projectile velocity; the skill and strength of the human using the weapons: all of these variables

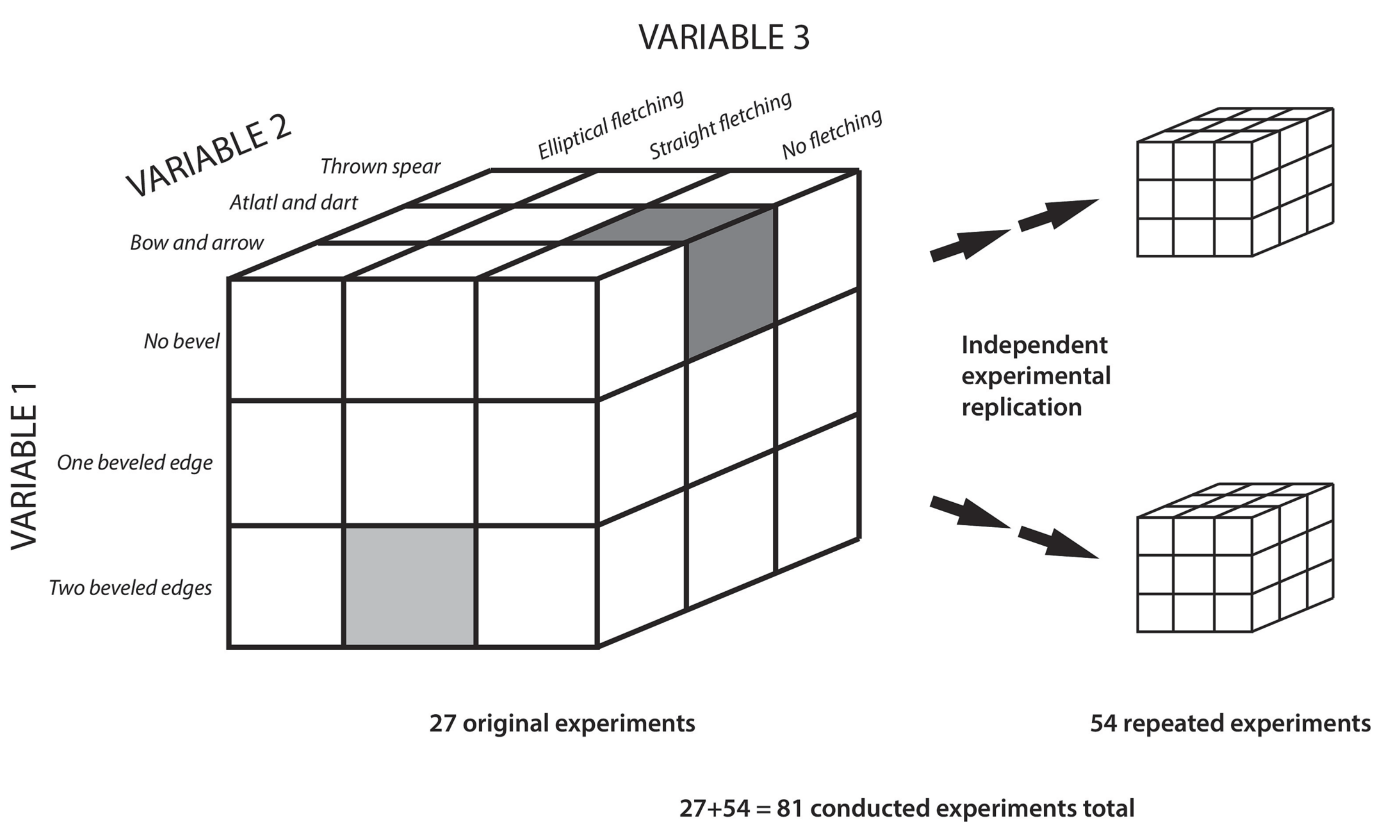

Fig. 2. A paradigmatic classification approach to experimental archaeology illustrates the vast number of possible experiments, and helps archaeologists understand any one experiment is merely a small step in understanding. See text for full explanation.

and more (including ones archaeologists may not be aware of) could potentially contribute to the degree of rotation in beveled versus non-beveled points. And all of these could be variables or variable states in our paradigmatic classification of experimental variables, adding hundreds of potential experiments that could be run. Several hundred more could be added to our paradigmatic classification by adding a variable with variable states that systematically alter the level of experimental control (e.g., amount of variable isolation or interaction).

And those thousands of experiments should not be conducted just once, but numerous times, and they would need to be repeated by others.

All that said, we are under no illusion that thousands of archaeological experiments will now be conducted on beveled points' rotation in flight by us or others. Given limited resources and time, the investigation of some experimental variables or variable combinations may take much higher priority over others. None of this is to say that experimental archaeologists should not use their results to make inferences about the past. But as this paradigmatic approach makes clear, any one archaeological experiment investigating a specific topic is merely a drop in the ocean of potential experiments. When framed in this way, it quickly becomes apparent that being too inferentially ambitious after only one or two experiments will not lead to meaningful results.

Another advantage of thinking about experimental variables via a paradigmatic approach is that it emphasizes that two experiments can both potentially be "right," even if their conclusions are in opposition. Consider, once again, our three variables, point type, weapon system, and fletching (Fig. 3**, left**). It is conceivable, for example, that the interacting variable states of "two beveled edges," "bow and arrow," and "elliptical fletching" could result in observable in-fight rotation, whereas the variable states of "two beveled edges," "atlatl and dart," and "elliptical fletching" would not. Moreover, experimental variable interaction thresholds might be identified that could potentially be compared to archaeological data sets of varying spatio-temporal scales (Fig. 3**, right**).

Thinking along these lines could be one way to help temper inferential over-ambition and over-confidence in experimental archaeology. There are two additional strategies we might also suggest that experimental archaeologists practice to help ensure they remain on firmer inferential ground. The first is to avoid the 'advocacy method' (*sensu* Wilson, 1975:28), whereby an author (or authors) proposes a hypothesis, selects or arranges evidence in the most persuasive manner possible, while another author (or authors) rebuts that hypothesis in whole or in part and raises an alternative argued with equal conviction. Such an approach is susceptible to unconscious bias, as well as the tendency to skate over issues of equifinality, as for example the myriad of (often) unknown ways in which past technologies were made, used, broken, or discarded.

A second strategy we suggest is that experimental archaeologists collaborate with non-experimentalists who are not enamored by the former's talent, be it in flintknapping, atlatl-throwing, or other ancient technologies or activities. Collaborating with archaeological colleagues who are indifferent to their craftsmanship or hunting prowess, and the checks and balance that should follow, will better insure the focus of an experiment is on testing hypotheses, rather than on imposing the authority of personal experiences onto interpretations of the archaeological record.

5. Conclusion

Experiments are a basic and vital part of the scientific process, and there is no reason archaeological science cannot benefit from them. However, while important strides in experimental archaeology have been made in recent years, there is still much work to be done. Here, we have used the disagreement over the function of beveled points to highlight one important issue experimental archaeologists face: balancing experimental results with the archaeological inferences ultimately drawn from them. We have argued that while the experimental contributions of both Lipo et al. (2012) and Pettigrew et al. (2015) are of value, they were each overly ambitious in their archaeological inferences. Their cases are scarcely unique in the realm of archaeological experimentation.

One way to close and strengthen the gap between archaeological experiments and archaeological inference is to frame experimental variables via a paradigmatic approach. Paradigmatic classification reveals the large number of potential experiments that can be conducted using different experimental conditions, which in turn highlights how little experimental archaeologists currently understand, and thus emphasizes the need for inferential restraint.

One can with an experimental approach far more reliably reject a claim than prove one, dependent of course on the specific conditions tested. Experiments thus move us closer to an answer, but because we are dealing with the past, we will never know with certainty (not even in regard to wholly rejecting a claim), and need to become comfortable with degrees of uncertainty. There are no quick and easy answers in experimental archaeology, nor shortcuts to robust archaeological inferences.

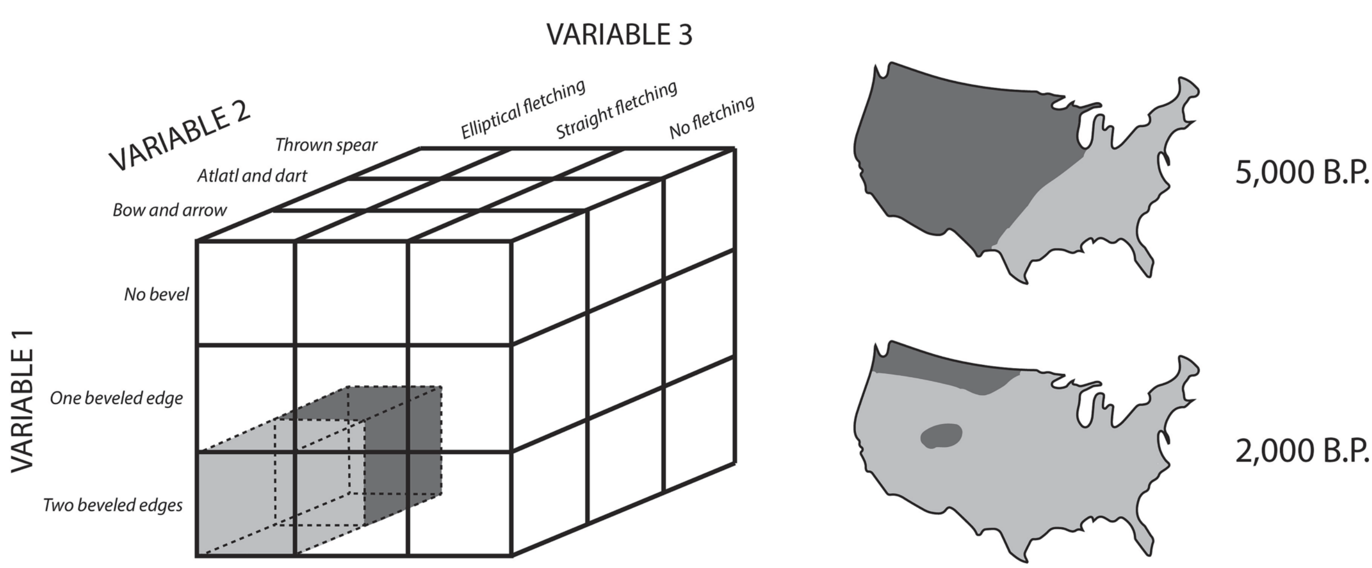

Fig. 3. Two different experiments (e.g. light gray and dark gray) may yield different results, but each may be applicable to different spatio-temporal scales.

It was experimental overreach that led to Binford's "ego-graphic analogy" and Thomas' "flintknapper's fundamental conceit." Even though experimental archaeology got a bad rap from these critiques, it was not wholly undeserved. Honesty compels the admission that their critiques are and should be ultimately beneficial to the maturation of experimental archaeology.

CRediT authorship contribution statement

Metin I. Eren: Conceptualization, Investigation, Writing – original draft, Writing – review & editing. **David J. Meltzer:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing.

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

No data was used for the research described in the article.

Acknowledgements

We thank Chris Hunt, the associate editor, and the two peer-reviewers for their constructive comments on this manuscript, which improved it in several ways. However, any mistakes are our own.

References

Ashby, E., 2007. Why single-bevel broadheads? Report from Ashby Bowhunting Foundation (https://www.ashbybowhunting.org/). Accessed 25 May 2023, https://static1.squarespace.com/static/5d0443b188b6c900011e0ccc/t/5d0e5bcbb3f1e700013faa8c/1561222093571/Single_Bevel_Broadheads.pdf.

Baldino, J., McKinny, S., Taylor, J., Wilson, M., Buchanan, B., Walker, R.S., Story, B.S., Bebber, M.R., Eren, M.I., 2023. North American Clovis Point Form and Performance V: An Experimental Assessment of Spear Thrusting Penetration Depth and Entry Wound Size. In Press, Lithic Technology.

Bebber, M.R., Buchanan, B., Eren, M.I., Walker, R.S., Zirkle, D., 2023. Atlatl use equalizes female and male projectile weapon velocity. Sci. Rep. 13 (1), 13349.

Bever, M. R., & Meltzer, D. J., 2007. Exploring variation in Paleoindian life ways: the third revised edition of the Texas Clovis fluted point survey. Bulletin of the Texas Archaeological Society, 78, 65-99.

Binford, L.R., 1981. Bones: Ancient Men and Modern Myths. Academic Press, New York.

Bonnichsen, R., 1983. Book review of *Bones: Ancient Men and Modern Myths*. Plains Anthropol. 28, 247–249.

Bradley, B. A., 1997. Sloan Site biface and projectile point technology. In *Sloan: A Paleoindian Dalton Cemetery in Arkansas*, by Morse, D. F., pp. 53-57. University of Arkansas Press, Fayetteville.

Calandra, I., Gneisinger, W., Marreiros, J., 2020. A versatile mechanized setup for controlled experiments in archeology. STAR: Science & Technology of Archaeological Research 6, 30–40.

Campbell, J.M., Ellis, F.H., 1952. The Atrisco Sites: Cochise Manifestations in the Middle Rio Grande Valley. Am. Antiq. 17, 211–221.

Coles, J., 1973. Archaeology by Experiment. Hutchinson, London.

Coles, J., 1979. Experimental Archaeology. Academic Press, London.

Conrad, G., Hough, S., Baldino, J., Gala, N., Buchanan, B., Walker, R.S., Key, A., Redmond, B., Bebber, M., Eren, M.I., 2023. Clovis bone versus stone weapon tip penetration: Thinking about relative costs and benefits, experimental assumptions, and archaeological unknowns at Sheriden Cave, Ohio, USA. J. Archaeol. Sci. Rep. 52, 104295.

Coppe, J., Lepers, C., Clarenne, V., Delaunois, E., Pirlot, M., Rots, V., 2019. Ballistic study tackles kinetic energy values of Palaeolithic weaponry. Archaeometry 61 (4), 933–956.

Dawkins, R., 1976. The Selfish Gene. Oxford University Press, Oxford.

Dibble, H.L., 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. J. Archaeol. Sci. 36, 1945–1954.

Dunnell, R.C., 1971. Systematics in Prehistory. The Free Press, New York.

Eren, M.I., Bebber, M.R., 2019. The Cerutti Mastodon site and experimental archaeology's quiet coming of age. Antiquity 93, 796–797.

Eren, M.I., Lycett, S.J., Patten, R.J., Buchanan, B., Pargeter, J., O'Brien, M.J., 2016a. Test, model, and method validation: the role of experimental stone artifact replication in hypothesis-driven archaeology. Ethnoarchaeology 8, 103–136.

Eren, M.I., Redmond, B.G., Miller, G.L., Buchanan, B., Boulanger, M.T., Hall, A., Hall, L., 2016b. The Wauseon Clovis fluted point preform, Northwest Ohio, USA: Observations, geometric morphometrics, microwear, and toolstone procurement distance. J. Archaeol. Sci. Rep. 10, 147–154.

Eren, M.I., Story, B., Perrone, A., Bebber, M., Hamilton, M., Walker, R., Buchanan, B., 2020. North American Clovis point form and performance: An experimental assessment of penetration depth. Lithic Technol. 45 (4), 263–282.

Eren, M.I., Mullen, D., Spurlock, L., Christy, B., Miller, G.L., Buchanan, B., Bush, J., Bebber, M.R., 2021a. Technical Descriptions of Artifacts from the Dresden Mound Cache, Ohio. Current Research in Ohio Archaeology 1–30.

Eren, M.I., Meltzer, D.J., Story, B., Buchanan, B., Yeager, D., Bebber, M.R., 2021b. On the efficacy of Clovis fluted points for hunting proboscideans. J. Archaeol. Sci. Rep. 39, 103166.

Eren, M.I., Bebber, M.R., Knell, E.J., Story, B., Buchanan, B., 2022a. Plains Paleoindian projectile point penetration potential. J. Anthropol. Res. 78, 84–112.

Eren, M.I., Meltzer, D.J., Story, B., Buchanan, B., Yeager, D., Bebber, M.R., 2022b. Not just for proboscidean hunting: On the efficacy and functions of Clovis fluted points. J. Archaeol. Sci. Rep. 45, 103601.

Eren, M.I., Bebber, M.R., Wilcox, D., Story, B., Buchanan, B., 2022c. North American Clovis point form and performance II: An experimental assessment of point, haft, and shaft durability. Lithic Technol. 47 (1), 38–51.

Eren, M.I., Miller, G.L., Story, B., Wilson, M., Bebber, M.R., Buchanan, B., 2023. North American Clovis Point Form and Performance IV: An Experimental Assessment of Knife Edge Effectiveness and Wear. In Press, Lithic Technology.

Eve, S., 2018. Losing our senses, an exploration of 3D object scanning. Open Archaeology 4 (1), 114–122.

Ferguson, J.R. (Ed.), 2010. Designing Experimental Research in Archaeology: Examining Technology through Production and Use. University Press of Colorado, Boulder.

Frison, G.C., 1989. Experimental use of Clovis weaponry and tools on African elephants. Am. Antiq. 54, 766–784.

Frison, G.C., 2004. Survival by Hunting: Prehistoric Human Predators and Animal Prey. University of California Press, Berkeley.

Goodyear, A. C., 1974. The Brand Site: A Techno-Functional Study of a Dalton Site in Northeast Arkansas (No. 7). Arkansas Archeological Survey, Fayetteville.

Harris, J.A., Boyd, R., Wood, B.M., 2021. The role of causal knowledge in the evolution of traditional technology. Curr. Biol. 31 (8), 1798–1803.

Hester, T.R., 1993. Lithic typology: background, goals, and a personal perspective. Lithic Technology 18 (1–2), 36–45.

Iovita, R., Schönekeß, H., Gaudzinski-Windheuser, S., Jäger, F., 2014. Projectile impact fractures and launching mechanisms: results of a controlled ballistic experiment using replica Levallois points. J. Archaeol. Sci. 48, 73–83.

Jennings, T.A., Smallwood, A.M., Pevny, C.D., 2021. Reviewing the role of experimentation in reconstructing Paleoamerican lithic technologies. PaleoAmerica 7, 53–67.

Key, A.J., 2016. Integrating mechanical and ergonomic research within functional and morphological analyses of lithic cutting technology: key principles and future experimental directions. Ethnoarchaeology 8, 69–89.

Key, A.J., Lycett, S.J., 2017a. Influence of handaxe size and shape on cutting efficiency: a large-scale experiment and morphometric analysis. J. Archaeol. Method Theory 24, 514–541.

Key, A.J., Lycett, S.J., 2017b. Reassessing the production of handaxes versus flakes from a functional perspective. Archaeological and Anthropological Sciences 9, 737–753.

Kilby, J.D., Surovell, T.A., Huckell, B.B., Ringstaff, C.W., Hamilton, M.J., Haynes Jr, C.V., 2022. Evidence supports the efficacy of Clovis points for hunting proboscideans. Journal of Archaeological Science: Reports 45, 103600.

Li, L., Lin, S. C., McPherron, S. P., Abdolahzadeh, A., Chan, A., Dogandžić, T., Iovita, R., Leader, G., Magnani, M., Rezek, Z., & Dibble, H. L., 2023. A Synthesis of the Dibble et al. Controlled Experiments into the Mechanics of Lithic Production. Journal of Archaeological Method and Theory, In Press.

Lin, S.C., Rezek, Z., Dibble, H.L., 2018. Experimental design and experimental inference in stone artifact archaeology. J. Archaeol. Method Theory 25, 663–688.

Lipo, C.P., Dunnell, R.C., O'Brien, M.J., Harper, V., Dudgeon, J., 2012. Beveled projectile points and ballistics technology. Am. Antiq. 77, 774–788.

Lohse, J.C., Collins, M.B., Bradley, B., 2014. Controlled overshot flaking: A response to Eren, Patten, O'Brien, and Meltzer. Lithic Technol. 39, 46–54.

Lycett, S.J., Chauhan, P.R., 2010. Analytical approaches to Palaeolithic technologies: an introduction. In: Lycett, S.J., Chauhan, P.R. (Eds.), New Perspectives on Old Stones: Analytical Approaches to Paleolithic Technologies. Springer, New York, pp. 1–22.

Lycett, S.J., Eren, M.I., 2013. Levallois lessons: the challenge of integrating mathematical models, quantitative experiments and the archaeological record. World Archaeol. 45 (4), 519–538.

Magnani, M., Rezek, Z., Lin, S.C., Chan, A., Dibble, H.L., 2014. Flake variation in relation to the application of force. J. Archaeol. Sci. 46, 37–49.

Magnani, M., Grindle, D., Loomis, S., Kim, A.M., Egbers, V., Clindaniel, J., Hartford, A., Johnson, E., Weber, S., Campbell, W., 2019a. Evaluating claims for an early peopling of the Americas: experimental design and the Cerutti Mastodon site. Antiquity 93, 789–795.

Magnani, M., Grindle, D., Loomis, S., Kim, A.M., Egbers, V., Clindaniel, J., Hartford, A., Johnson, E., Weber, S., Campbell, W., 2019b. Experimental futures in archaeology. Antiquity 93, 808–810.

Martellotta, E.F., Perston, Y.L., Craft, P., Wilkins, J., Langley, M.C., 2022. Beyond the main function: An experimental study of the use of hardwood boomerangs in retouching activities. PLoSONE 17 (8), e0273118.

Meltzer, D.J., 2015. The Great Paleolithic War. University of Chicago Press, Chicago.

Mesoudi, A., 2011. Cultural Evolution. University of Chicago Press, Chicago.

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

Mika, A., Buchanan, B., Walker, R., Key, A., Story, B., Bebber, M., Eren, M.I., 2022. North American Clovis point form and performance III: An experimental assessment of knife cutting efficiency. Lithic Technol. 47 (3), 203–220.

Milks, A., 2019. Skills shortage: A critical evaluation of the use of human participants in early spear experiments. EXARC Journal 2019 (2), 1–11.

Milks, A., Champion, S., Cowper, E., Pope, M., Carr, D., 2016. Early spears as thrusting weapons: Isolating force and impact velocities in human performance trials. J. Archaeol. Sci. Rep. 10, 191–203.

Milks, A., Parker, D., Pope, M., 2019. External ballistics of Pleistocene hand-thrown spears: experimental performance data and implications for human evolution. Sci. Rep. 9 (1), 820.

Mraz, V., Fisch, M., Eren, M.I., Lovejoy, C.O., Buchanan, B., 2019. Thermal engineering of stone increased prehistoric toolmaking skill. Sci. Rep. 9 (1), 14591.

Mullen, D., Sitton, J., Story, B., Buchanan, B., Walker, R.S., Eren, M.I., Bebber, M.R., 2023. Comparison of four ballistic and thrusting target materials: an experimental and Bayesian approach using static testing of stone and steel arrow tips. Archaeometry 65, 1108–1124.

Muller, A., 2017. The role of experimental knapping in empirically testing key themes in the evolution of lithic technology: reduction intensity, efficiency and behavioural complexity. University of Queensland, Australia. Ph.D. Dissertation,.

Muller, A., Shipton, C., Clarkson, C., 2022. Stone toolmaking difficulty and the evolution of hominin technological skills. Sci. Rep. 12 (1), 5883.

Neill, L., Clarkson, C., Schoville, B., 2022. Holding your shape: Controlled tip fracture experiments on cast porcelain points. J. Archaeol. Sci. Rep. 44, 103505.

Outram, A.K., 2008. Introduction to experimental archaeology. World Archaeol. 40, 1–6.

Pargeter, J., Brooks, A., Douze, K., Eren, M.I., Groucutt, H.S., McNeil, J., Mackay, A., Ranhorn, K., Scerri, E., Shaw, M., Tryon, C., Will, M., Leplongeon, A., 2023a. Replicability in Lithic Analysis. Am. Antiq. 88 (2), 163–186.

Pargeter, J., Liu, C., Kilgore, M.B., Majoe, A., Stout, D., 2023b. Testing the effect of learning conditions and individual motor/cognitive differences on knapping skill acquisition. J. Archaeol. Method Theory 30 (1), 127–171.

Patten, R.J., 1981. Review of Hardaker's taxonomy. *Flintknapper's*. Exchange 4, 12.

Perreault, C., 2020. The Quality of the Archaeological Record. University of Chicago Press, Chicago.

Pettigrew, D.B., Whittaker, J.C., Garnett, J., Hashman, P., 2015. How atlatl darts behave: beveled points and the relevance of controlled experiments. Am. Antiq. 80, 590–601.

Rezek, Z., Lin, S.C., Dibble, H.L., 2016. The role of controlled experiments in understanding variation in flake production. In: Sullivan, A.P., Olszewski, D.I. (Eds.), Archaeological Variability and Interpretation in Global Perspective. University Press of Colorado, Boulder, pp. 307–320.

Schiffer, M.B., 2016. Behavioral Archaeology: Principles and Practice. Routledge, New York.

Schiffer, M. B., 2013. Contributions of experimental archaeology. In *The Archaeology of Science: Studying the Creation of Useful Knowledge*, by Schiffer, M. B., pp. 43-52. Springer, New York.

Schillinger, K., Mesoudi, A., Lycett, S.J., 2014. Copying error and the cultural evolution of "additive" vs. "reductive" material traditions: an experimental assessment. Am. Antiq. 79, 128–143.

Schillinger, K., Mesoudi, A., Lycett, S.J., 2015. The impact of imitative versus emulative learning mechanisms on artifactual variation: implications for the evolution of material culture. Evol. Hum. Behav. 36, 446–455.

Shea, J.J., 2020. Prehistoric Stone Tools of Eastern Africa: A Guide. Cambridge University Press, Cambridge.

Sisk, M.L., Shea, J.J., 2009. Experimental use and quantitative performance analysis of triangular flakes (Levallois points) used as arrowheads. J. Archaeol. Sci. 36 (9), 2039–2047.

Smallwood, A.M., Pevny, C.D., Jennings, T.A., Morrow, J.E., 2020. Projectile? Knife? Perforator? Using actualistic experiments to build models for identifying microscopic usewear traces on Dalton points from the Brand site, Arkansas, North America. Journal of Archaeological Science: Reports 31, 102337.

Smith, A.G., 1953. Beveled or "Rotary" Points. Am. Antiq. 18, 269–270.

Sollberger, J.B., 1971. A technological study of beveled knives. The Plains Anthropologist 16, 209–218.

Stanford, D., Bradley, B.A., 2012. Across Atlantic Ice. University of California Press, Berkeley.

Stemp, W.J., 2016. Twist and shout: Experiments in ancient Maya blood-letting by piercing with obsidian blades and splinters. J. Archaeol. Sci. Rep. 9, 134–142.

Stenak, S., 2022. Confined Pressure Compression Scratches: A New Observation in Lithic Analysis. Lithic Technol. 47 (1), 13–22.

Story, B.A., Eren, M.I., Thomas, K., Buchanan, B., Meltzer, D.J., 2019. Why Are Clovis Fluted Points More Resilient than Non-Fluted Lanceolate Points? A Quantitative Assessment of Breakage Patterns Between Experimental Models. Archaeometry 61, 1–13.

Thomas, D.H., 1986. Points on points: A reply to Flenniken and Raymond. Am. Antiq. 51 (3), 619–627.

Thomas, K.A., Story, B.A., Eren, M.I., Buchanan, B., Andrews, B.N., O'Brien, M.J., Meltzer, D.J., 2017. Explaining the origin of fluting in North American Pleistocene weaponry. J. Archaeol. Sci. 81, 23–30.

Wilkins, J., Schoville, B.J., Brown, K.S., Chazan, M., 2012. Evidence for early hafted hunting technology. Science 338 (6109), 942–946.

Wilson, T., 1898. Beveled Arrowheads. American Archaeologist 2, 141–143.

Wilson, E.O., 1975. Sociobiology: the new synthesis. Belknap Press, Harvard.

Wilson, T., 1899. Arrowpoints, Spearheads, and Knives of Prehistoric Times. Report of the United States National Museum 1897:811-988. Washington, D.C.

Wobst, H.M., 1978. The archaeo-ethnology of hunter-gatherers or the Tyranny of the Ethnographic Record in Archaeology. American Antiquity 43, 303–309.