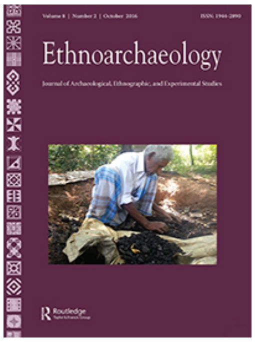

Routledge
Taylor & Francis Group

Ethnoarchaeology
Journal of Archaeological, Ethnographic and Experimental Studies

ISSN: 1944-2890 (Print) 1944-2904 (Online) Journal homepage: http://www.tandfonline.com/loi/yeth20

Test, Model, and Method Validation: The Role of Experimental Stone Artifact Replication in Hypothesis-driven Archaeology

Metin I. Eren, Stephen J. Lycett, Robert J. Patten, Briggs Buchanan, Justin Pargeter & Michael J. O'Brien

To cite this article: Metin I. Eren, Stephen J. Lycett, Robert J. Patten, Briggs Buchanan, Justin Pargeter & Michael J. O'Brien (2016) Test, Model, and Method Validation: The Role of Experimental Stone Artifact Replication in Hypothesis-driven Archaeology, Ethnoarchaeology, 8:2, 103-136, DOI: 10.1080/19442890.2016.1213972

To link to this article: http://dx.doi.org/10.1080/19442890.2016.1213972

ETHNOARCHAEOLOGY, Vol. 8 No. 2, October, 2016, 103–136

Test, Model, and Method Validation: The Role of Experimental Stone Artifact Replication in Hypothesis-driven Archaeology

METIN I. EREN[1,2,3], STEPHEN J. LYCETT[4], ROBERT J. PATTEN[5], BRIGGS BUCHANAN[6], JUSTIN PARGETER[7,8] and MICHAEL J. O'BRIEN[9]

[1]Department of Anthropology, Kent State University, Kent, OH 44242, USA; [2]Department of Anthropology, University of Missouri, Columbia, MO 65211, USA; [3]Department of Archaeology, Cleveland Museum of Natural History, Cleveland, OH 44107, USA; [4]Department of Anthropology, University at Buffalo, SUNY, Buffalo, NY 14261, USA; [5]Stone Dagger Publications, Lakewood, CO 80232, USA; [6]Department of Anthropology, University of Tulsa, Tulsa, OK 74104, USA; [7]Department of Anthropology, Stony Brook University, Stony Brook, NY 11794-4364, USA; [8]Department of Anthropology and Development Studies at the University of Johannesburg, Auckland Park, South Africa; [9]Department of Anthropology, University of Missouri, Columbia, MO USA

For many years, intuition and common sense often guided the transference of patterning ostensibly evident in experimental flintknapping results to interpretations of the archaeological record, with little emphasis placed on hypothesis testing, experimental variables, experimental design, or statistical analysis of data. Today, archaeologists routinely take steps to address these issues. We build on these modern efforts by reviewing several important uses of replication experiments: (1) as a means of testing a question, hypothesis, or assumption about certain parameters of stone-tool technology; (2) as a model, in which information from empirically documented situations is used to generate predictions; and (3) as a means of validating analytical methods. This review highlights the important strategic role that stone artifact replication experiments must continue to play in further developing a scientific approach to archaeology.

DOI 10.1080/19442890.2016.1213972

KEYWORDS experiment, flintknapping, lithic technology, method validation, model, replication, stone tools, test

In 1978, L. Lewis Johnson published a history of flintknapping experimentation that covered the period 1838–1976 (Johnson 1978). The lengthy article appeared in *Current Anthropology*, a journal whose format allows for companion commentaries by experts in a particular field. Those that accompanied Johnson's article were anything but complimentary of the then-current state of flintknapping experimentation. For example, Cahen (1978, 360) noted, "I am afraid that stone-knapping experimentation tends to be an achievement in itself instead of an approach to the reality of a prehistoric way of life," and Hay (1978, 361) commented, "general anthropological interests are completely lost in a multitude of issues and problems specific to flint knapping itself." Malik (1978, 364) was even more negative: "apart from the chronological compilation of the history of lithic experimentation I fail to see what contribution it makes to the understanding of lithic tool technology." Katz (1978, 362) derisively advised "it's time to look seriously at our models instead of admiring them." Perhaps most damning was the comment by Müller-Beck (1978, 364) that "nearly all the experiments are empirical ones… they are – so far – not really controlled experiments"—a point Hayashi (1968) had made a decade earlier. In other words, no theory guides why one experiment is performed as opposed to an alternative (see also Dincauze 1978; Knudson 1978; Ranere 1978).

One might have thought that a set of comments that critical, especially appearing as they did in a leading anthropological journal, would have led to changes in how flintkapping experiments were designed and carried out (such as those outlined in Clarke 1968; Dunnell 1971; Eren et al. 2014a; Lett 1997; Lycett and Chauhan 2010; O'Brien 2010; Surovell 2009), but with a few exceptions, archaeologists, especially those who were expert flintknappers, continued to use the craft to make authoritative, intuitive arguments about lithic technology. This tendency was identified in a critique by Thomas (1986:623), who called it "the flintknapper's fundamental conceit":

> At the heart of the matter is the vexing conceit that underlies too much of contemporary lithic technology: some flintknappers behave as if the act of breaking rocks gives them an inside track to the truth. This attitude is reflected throughout the work of Flenniken (especially 1984; Flenniken and Raymond 1986), but he is hardly alone.

Thomas was correct: Although he singled out Flenniken's work for criticism, the same could have been said about much of the lithic analysis that had been done up to that time. In many respects, it had grown up as a cottage industry, with practitioners imposing interpretations about prehistoric human behavior on the archaeological record based on their own intuition, impressions, and mastery of the flintknapping craft. Lost was the fact that "commonly accepted" or "proposed by an experienced knapper" were not the same thing as empirical support within an explicit hypothesis-testing framework. We accept that hard-won expertise in the craft of flintknapping might indeed provide a route to a more informed opinion than one proffered in the absence of practical experience and awareness. Nevertheless, the danger of an authoritative stance being used in place of more formal scientific procedures and rigorous analysis was clearly beginning to raise major concerns at this time.

By the late 1980s and 1990s the tide began to move away from intuitive analysis and toward the use of replicated stone tools to test hypotheses about lithic technology, to construct models against which archaeological data could be compared, and to evaluate methods that could be applied to archaeological data (e.g. Barham 1987; Bradley and Sampson 1986; Davis and Shea 1998; Kuhn 1990; Odell and Cowan 1986; Roux, Bril, and Dietrich 1995; Whittaker 1987). Within a decade or so, a number of controlled replication experiments appeared (e.g. Clarkson 2002; Shott et al. 2000; Stout et al. 2000; Whittaker and McCall 2001), signaling the emergence of replication as an important source of information regarding prehistoric human behavior.

Our objective here is not to provide encyclopedic coverage of modern stone-tool replication but rather to examine some of the theoretical and design-related underpinnings of modern efforts in terms of replication as test, as model, and as method validation. Other categorizations of replication experiments exist (e.g. Carr and Bradbury 2010; Flenniken 1984; Lerner 2013; Nami 2010; Olausson 2010; Shea 2015; Whittaker 1994), but we find these three categories useful for organizing the growing literature on the subject and highlighting how specific experiments have contributed to a better understanding of human behavior. The review hopefully will be of interest to archaeologists—flintkappers as well as non-flintknappers—who want to construct hypotheses about prehistoric tool manufacture and derive testable implications from them.

What Is Stone-tool Replication?

Stone-tool replication can be defined as the act of creating or using non-artifactual flaked-stone specimens for the purpose of investigating archaeological hypotheses, questions, and methods. In some circumstances it may be useful to differentiate between stone-tool replication and "flintknapping," which Reti (2014) defines as the act of creating flaked-stone tools as art, for personal pleasure (a hobby), or for business purposes (e.g. eighteenth-century gun flints) (see also Whittaker 2004). While a "replica" is often thought of as an object that matches as closely as possible to a specific original, a broader definition of the term "replica" may be more useful for our purposes here, namely a new-made object that possesses attributes relevant to better understanding prehistoric artifacts. Stone-tool replication falls generally under the subheading of experimental archaeology, although equating it with "experiment" would be misleading. If we define "experiment" as a form of scientific study that uses a structured, replicable procedure to test the validity of a hypothesis (Outram 2008), then it should be clear that stone-tool replication is but a single part of a process that also includes construction of hypotheses, derivation of test implications, and use of appropriate analytical methods (Lycett and

Chauhan 2010). Equating stone-tool replication with "experiment" would be like equating the act of pouring chemicals into a beaker with a chemistry "experiment." Stone-tool replication should be considered to be an act or task that is part of the experimental process, but not that process itself.

Replication experiments form an important link in a chain of archaeological investigations ranging from studies of the archaeological record itself to mathematical models of that record (Lycett and Eren 2013a). This chain is parallel to another that spans a theoretical range of increasing "external validity" toward one end, and increasing "internal validity" toward the other (Mesoudi 2011; Roe and Just 2009). In archaeology, the artifactual record might be considered to have high "external validity": it is the most direct, empirical (i.e., tangible) evidence that we have of what actually took place in the past. Conversely, however, the artifactual record has low "internal validity": it is inherently biased, incomplete, and allows little control or randomization of the variables it produces. Further, excavation is unrepeatable and if one variable/trend/pattern is observed it may be difficult to determine how "typical" this may have been in other regions or temporal spans where such evidence is not currently available. In direct contrast to this, experiments might be considered to have high "internal validity": they can be repeated, and their parameters and variables might be controlled and manipulated in multiple ways (Mesoudi 2011, 135; see also discussion in Clarkson, Haslam, and Harris 2015a, 121; Pettigrew et al. 2015). There is, however, an inevitable cost to this high internal validity: no experiment can "re-run" prehistory with exact precision, and its relationship to the parameters of direct interest (i.e., the archaeological record) requires the imposition of specific assumptions and inferences to give it archaeological meaning. The archaeological record, despite its problems, is the best and only direct evidence of the past that we have; experiments are very much an indirect means of making inferences regarding the past. In this sense, "internal validity" and "external validity" can be seen to refer to opposing strengths and weaknesses in reference to the data provided directly by the archaeological record versus experiments (Lycett and Eren 2013a).[1]

The issue of external versus internal validity leads to the related issue of replication experiments that involve the use of machines or other devices versus those that involve human knappers. Machine flaking has provided valuable insights into some potential causal variables that govern stone-tool fracture mechanics at the level of individual flakes[2] (e.g. Dibble 1997, 1998; Dibble and Pelcin 1995; Dibble and Rezek 2009; Dibble and Whittaker 1981; Magnani et al. 2014; Pelcin 1997a, 1997b, 1997c, 1998; Rezek et al. 2011), but it would be a mistake to assume machine-flaking experiments are automatically superior to human ones or vice versa. Certain variables such as "force of blow" or "angle of blow" can be measured or observed more easily through machine flaking than through human flaking, but the design of the machine itself may introduce variables whose effects on stone fracture relative to what is present in the archaeological record are unclear. Likewise, "control" may actually remove important interactions between variables that were operable in the past and so lead to distorted results with respect to understanding archaeological questions. Thus, experimental control is a *strategy* in which any perceived benefit from one degree or kind of control necessarily comes with an unavoidable cost. Not acknowledging these points and consistently advocating for greater and greater experimental control might be considered a controlled experimenter's fundamental conceit.

Given these issues, a pragmatic rather than dogmatic approach to the role of experiments in archaeological inference is needed, especially given that the past can never be truly "replicated." Suffice it to say that formal experiments involving machines on the one hand and human flintknappers on the other are not in opposition; rather, they occupy different spaces on the spectrum of analytical validity (Figure 1). Increased linkage between these analytical spaces may arise in future years given advances in technology and digital imaging. Accurate measurements of force or angle of blow or other mechanical and biomechanical variables may soon be recorded in investigations of human flintknappers, while advances in robotics may allow machine flaking to be conducted in a more human-like fashion. Until then, however, no machine can knap a replica Clovis point or Levallois core, so there are some questions or topics of inquiry that currently only human stone-tool replication can investigate. Human stone-tool replication, therefore, has the capacity to provide an important bridge between highly abstract, mathematical models or experiments where variables have been controlled in a very artificial manner, and the archaeological record (Lycett and Eren 2013a).

An analogous point could be made about reductive experiments that use humans but not stone or other materials with conchoidal fracture (e.g. glass). Schillinger, Mesoudi, and Lycett (2014a, 2014b, 2015), for example, used standardized blocks of foam and plasticine as a substitute for stone, and plastic knives as a substitute for hammerstones and antler billets, to gain insights into such variables as copying error, additive versus reductive processes, time budgets, and imitative versus emulative learning mechanisms, which are all relevant in the context of stone-tool manufacture. Foam and plasticine do not flake or reduce in the same way as material types possessing conchoidal fracture, yet despite this artificiality, Schillinger

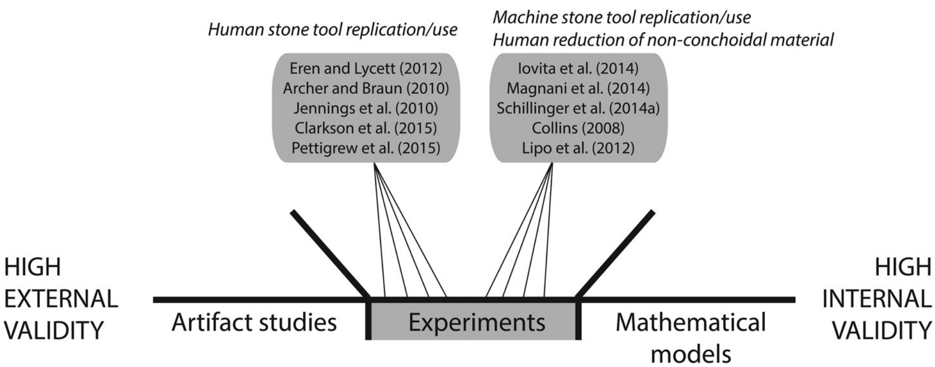

FIGURE 1 Experiments that use human flintknappers versus those that use machines, or those in which a reductive material other than flaked stone is used, are not opposed, but merely occupy different spaces on the same analytical spectrum: the former generally occupy a space with higher external, but lower internal, validity while the latter generally occupy a space with higher internal, but lower external, validity.

and colleagues' experiments contribute to our understanding of flaked stone because of their ability to directly address parameters of interest in the context of stone tools. Additionally, the lack of any actual stone-tool replication means that virtually anyone can participate, and hence large sample sizes, and statistical robustness with respect to those parameters of interest, are easily achieved. Indeed, such studies illustrate the potential for artifact-focused experiments to create bridges to other questions and bodies of literature beyond lithic analyses and archaeologically specific endeavors alone.

Designing a Replication Experiment

For stone-tool replication to be useful, there first must exist a hypothesis or question from which can be drawn clear, empirical predictions. Like in any scientific study, the hypothesis and its predictions determine the variables required for the experiment. Experimental variables include such things as the sample size of participants or specimens, the measurement and test protocols, whether the experiment is a blind test (see below), and the chosen quantitative methods and statistical analyses. Additional experimental variables specific to replication include the reduction strategy, skill-level of the knapper, material type, number and types of knapping tools available, and perhaps how the knapper is situated (e.g. sitting in a chair versus on the ground) (Carr and Bradbury 2010). In some instances it might not matter whether the raw material is a basalt from Africa or a chert from Texas, whether the knapper is skilled or not, or whether soft- or hard-hammer percussion is used. The design of an experiment and the variables that go into it must be considered carefully to understand what matters and what does not in the context of a specific question, and what could thus validate or confound the results of an experiment. Whatever variables and test procedures go into an experimental design, they all must be recorded carefully so that they can be described explicitly and in full when the time comes to publish. When possible, quantification should be utilized over qualitative description since this improves precision, comparability, and evaluation of results.

This discussion might sound axiomatic or commonsensical, but we have encountered archaeologists who think that the mere act of "busting rocks" or using a stone tool to butcher an animal constitutes publishable research. This may have been the case at one time in the same way that the act of dissecting a mollusk would have resulted in a published biology paper 150 years ago (e.g. Owen 1835). These kinds of exploratory or experiential studies have value and are one possible means of generating new hypotheses, questions, and ideas[3] (McCall and Pelton 2010; Shelley 1990), but as archaeology matures as a science, so too must the hypotheses posed and the questions asked, which in turn requires greater sophistication in experimental design, instrumentation, and recorded variables. By "greater sophistication" we do not necessarily mean increased use of state-of-the-art laboratory gadgets or computer software alone, but instead progressively better linkage between a hypothesis and its predictions with an experiment and its results. This kind of structured research requires organization, planning, forethought, and, quite often, pilot experiments. If in order to robustly test the predictions of a particular hypothesis the highest-resolution 3D scanner available on the market is needed to record attributes on replicated stone tools, and the knapper must be in a highly controlled and sterile laboratory, then so be it. But other hypotheses may actually be better tested in less-controlled experimental conditions in the field using simple calipers to record, for example, flake length, width, and thickness. It all depends on the hypothesis being tested. Many hypotheses would benefit from multiple replications that systematically modify relevant factors and juxtapose the results (Eren et al. 2011a:2014; Lycett and Eren 2013a; Mesoudi 2011; see also Carr and Bradbury 2010; Marsh and Ferguson 2010; Pettigrew et al. 2015).

Replication in Hypothesis-driven Archaeology

Recall our earlier brief discussion of three principle arenas in which experimental stone-tool replication contributes to a hypothesis-driven archaeology: as test, as model, and as method validation. As we pointed out, our list is not meant to imply that there are not other ways to categorize replication experiments; the categories are simply meant to act as archetypes with which to organize the ever-expanding literature in order to more clearly understand how specific experiments contribute to an understanding of hominin behavior. In reality, stone-tool replication experiments can, and often do, fall into two or more of these categories, though at times they may lean more heavily towards one versus another.

Replication as Test

In this category of experiment, stone-tool replication is used to test a question, a hypothesis, or an assumption about certain parameters of lithic technology (e.g. Lycett and Chauhan 2010:5). Results of such tests are not necessarily meant to be directly compared to archaeological data but instead to serve as a means of formally assessing and understanding the bounds of what is practically achievable when making or using stone tools in order to support or falsify potential motivating factors underlying patterns of tool production, use, morphology, and variability (Diez-Martin and Eren 2012; Lycett and Eren 2013b). There are several broad avenues of inquiry that have been investigated by means of experimental tests, including comparative morphology (Driscoll 2011; Eren and Lycett 2012; Gurtov, Buchanan, and Eren 2015; Presnyakova et al. 2015; Williams and Andrefsky 2011); process controls (Patten 2002, 2005, 2009); tool use-life (Shott 2002); cognition and language (Geribas, Mosquera, and Vergès 2010; Mahaney 2014; Morgan et al. 2015b; Putt, Woods, and Franciscus 2014; Stout et al. 2000; Uomini and Meyer 2013); biomechanics (Faisal et al. 2010; Key and Lycett 2011; Key and Dunmore 2015; Nonaka, Bril, and Rein 2010; Rolian, Lieberman, and Zermeno 2011; Williams, Gordon, and Richmond 2012, 2014); and the influence of stone raw material differences on lithic form (Archer and Braun 2010; Eren et al. 2014b), production technology (Bar-Yosef et al. 2012), tool function (Braun et al. 2009; Galán and Domínguez-Rodrigo 2014; Rodríguez-Rellán, Valcarce, and Esnaola 2013; Waguespack et al. 2009; Wilkins, Schoville, and Brown 2014), knapper skill (Duke and Pargeter 2015; Eren, Bradley, and Sampson 2011b, Eren et al. 2011c; Stout and Semaw 2006; Winton 2005), use-wear accrual (Lerner et al. 2007), and impact fractures (Pargeter 2013). Two additional avenues of inquiry—core-reduction efficiency and functional morphology—are discussed below.

Core-reduction efficiency

Core-reduction efficiency, measured in terms of total cutting edge produced, number of flakes produced, or time required for core reduction, has been investigated through several experimental tests (e.g. bipolar reduction: Diez-Martin et al. 2011, Li 2015, Morgan et al. 2015a; blade versus discoidal reduction: Eren, Greenspan, and Sampson 2008; Levallois reduction: Lycett and Eren 2013b; biface versus amorphous-core reduction: Prasciunas 2007; biface versus blade reduction: Rasic and Andrefsky 2001; see also Putt 2015; Tactikos 2003). Efficiency is notoriously difficult, if not impossible, to quantify from archaeological specimens for two reasons. First, entire reduction sequences must be present in order to calculate the original unmodified nodule mass as well as the total number and mass of all knapped flakes. Second, there is no way of knowing whether the knapper intended to reduce a nodule as efficiently as possible, or whether the knapper even possessed the skill to do so (see Duke and Pargeter 2015; Morgan et al. 2015a). Thus, to understand efficiency we turn to replication experiments in which an expert knapper can be instructed to knap cores as efficiently as possible, which controls for both skill and intention (Eren et al. 2011c).

The logic underlying experimental comparisons of core-reduction efficiency is that if strategy A is more efficient than strategy B for producing cutting edges or flakes, then reduction efficiency is a potential motivating factor for adopting strategy A over strategy B. If there is archaeological evidence that strategy B was adopted despite lower efficiency, then another explanation can be sought. Perhaps knappers did not possess the knowledge or skill to use strategy A, or perhaps strategy B provided other benefits that in certain contexts were more desirable than reduction efficiency, such as specific flake shapes. If, however, no significant difference is found through replication, then efficiency should be considered to be a nonfactor. Core-reduction efficiency can also be assessed within a single reduction strategy in order to assess the influence of particular variables on efficiency, such as knapper skill (Eren, Bradley, and Sampson 2011b) or stone raw material (Gurtov and Eren 2014).

A recent example of a replication experiment that tested core-reduction efficiency is that of Jennings, Pevny, and Dickens (2010), who compared the number of flake blanks produced and the transport mass of bifacial cores and wedge-type blade cores typical of Clovis Paleoindians in the western United States versus discoidal cores and amorphous cores typical of chronologically subsequent Folsom Paleoindians. While Jennings, Pevny, and Dickens (2010) investigated six replicated biface cores and five new blade cores, they bolstered their own experimental data with core-efficiency data from Prasciunas (2007, 10 biface cores and 10 amorphous cores) and Eren, Greenspan, and Sampson (2008, seven prismatic-type blade cores and seven discoid cores). They came to four conclusions:

1. When cores are small, amorphous cores are more efficient, but as cores increase in size, bifacial, discoidal, and blade cores approach amorphous cores in terms of production efficiency.
2. Small bifacial cores are less efficient than larger ones in terms of transport mass because larger ones produce more mass-efficient flakes.
3. Prismatic and wedge-type blade cores are equally efficient at producing flake blanks, and for both types efficiency decreases with core size.
4. Tentatively, bifacial reduction may be more efficient at producing noncortical flake blanks than blade reduction from wedge-type cores.

These conclusions allowed Jennings, Pevny, and Dickens (2010) to make several inferences about intra- and intercultural patterns of Clovis and Folsom tool making. For example, core size varies between the North American Southern Plains on the one hand and the Northern Plains and Rocky Mountains on the other. The former region contains numerous large tabular chert-nodule outcrops, whereas the latter two regions contain fewer outcrops, making small nodule and cobble sources more important. Based on their experimental results, Jennings, Pevny, and Dickens (2010) predicted different core-reduction strategies for Clovis and Folsom groups, namely the use of any or all reduction strategies in the Southern Plains and a relatively higher use of amorphous-core and discoidal-core reduction in the Rocky Mountains and on the Northern Plains. Informal assessment of the archaeological record suggested that Clovis knappers used biface and blade reduction, regardless of region, and that amorphous-core reduction was never dominant. Folsom knappers, however, conformed to the predictions of the experimental core-efficiency results, using biface cores on the Southern Plains and amorphous and discoidal cores in the other regions. Taken together, the experimental core-efficiency results and archaeological patterns inspired new, interesting, and empirically based interpretations of Paleoindian mobility, settlement, landscape use, and technological evolution.

Jennings, Pevny, and Dickens (2010) study raises several important points relevant to the process of replication experiments. First, Jennings and colleagues operated under a clear theoretical framework, with questions that logically preceded, and inspired, the experiment. This ensured that either a significant *or* non-significant result would have been of importance and interest. The experiment had a clear purpose, and the analysts recorded only those variables—flake-blank counts and flake mass—needed to address the issue at hand. No time, energy, or space was wasted on aimless data-mining in the hopes of finding a significant result. Second, Jennings and colleagues were explicit about such variables as who the knapper was, tools used, measurements of the original core nodule, how data were generated, (e.g. which flakes were included in the study), and which statistical tests were used. They also included a discussion of their experiment's limitations. Thus, independent re-testing becomes much more straightforward.

Jennings, Pevny, and Dickens (2010, 2157) specified the use of copper billets, which brings up a third point worth emphasizing. Paleoindians obviously did not use copper billets and industrial grinding stones—a fact that might lead some lithic analysts to claim that the core-efficiency results are invalid. This stance is not necessarily justified. In some cases, depending on the question posed, the employment of an experimental variable, such as a copper billet may indeed invalidate an experiment's results, inferences, or conclusions. In other cases, the choice of experimental variables such as knapping-tool material may have relatively little bearing on the specific phenomenon or archaeologically relevant variable toward which the overall experimental design is being strategically directed. Use of a copper billet could, in this case, be argued to increase Jennings, Pevny, and Dickens (2010) experiment's internal validity and consistency in that it controls for a variable that otherwise might inconsistently influence core efficiency in its own way if, say, antler or wood is used for blade reduction but a hard hammerstone is used for discoidal-core reduction. A series of well-controlled, blind stone-tool replication experiments systematically looking at the influence of copper billets and pressure tools versus antler, wood, and stone percussors and pressure tools on core-reduction efficiency, pattern/choice of flake removal, debitage flake morphology, final tool morphology, among other topics, would be extremely valuable. Once the results of these experiments are in hand, replication experimenters will be armed with information that will help them better design and interpret their experiments, namely when it would be most beneficial to employ copper or "natural" tools to best answer a specific question.

Functional morphology

Functional morphology is the study of the relationship between form (size and shape) and application toward specific tasks. Examples of replication experiments conducted to better understand the relationship between shape and use include Collins' (2008) study of the performance of differently shaped flake edges; Shea, Davis, and Brown's (2001) and Sisk and Shea's (2009) experimental tests of triangular flakes as arrowheads; studies of projectile point performance by Titmus and Woods (1986), Odell and Cowan (1986), Friis-Hansen (1990), Cheshier and Kelly (2006), and Hunzicker (2008); Braun, Pobiner, and Thompson's (2008) study of cutmark production, butchery activity, and tool edge attrition; Quinn et al.'s (2009) examination of the perforation capabilities of Pre-Pottery Neolithic el-Khiam points; Pétillon et al.'s (2011) study of functional characteristics of Magdalenian composite projectile tips; Eren et al.'s (2013) test of overshot versus overface flake thinning effectiveness; Key and Lycett's (2014; see also Prasciunas 2007) examination of flake size versus cutting efficiency; Key and Lycett's (2015) assessment of flake edge angle vs. cutting efficiency; Clarkson, Haslam, and Harris' (2015a) experimental trials of retouched, non-retouched, and hafted flake woodworking; and Lipo et al.'s (2012) and Pettigrew et al.'s (2015) examinations of projectile point beveling as a spinning and stabilizing mechanism. None of these studies was concerned primarily with comparing experimental results directly with archaeological specimens so much as using results to establish functional parameters that could help hone archaeological hypotheses and interpretations about potential functional options or limits of archaeological specimens. Since archaeologists cannot observe prehistoric hominins using their stone tools, much less watch hominins push their stone tools to some sort of maximum functional "ceiling," and there are ethical issues to consider regarding the modern experimental use of prehistoric artifacts (Chazan 2013), archaeologists must instead rely on modern experimental tests utilizing stone-tool replicas to establish boundaries of functional morphology.

One early example of replication used in a test of functional morphology was that of Frison (1989), who asked whether replicated Clovis points could penetrate tough, thick hides like those of modern elephants when delivered by atlatls or thrusting spears. Frison was testing the basic and widely held assumption that the occasional association of prehistoric Clovis points and mammoth remains was a result of Clovis people using the former to hunt the latter. If experimental tests using replicated Clovis points were found to successfully penetrate elephant hides, the argument for prehistoric Clovis points as mammoth-hunting implements would remain a reasonable one. If, on the other hand, replica Clovis points could not penetrate elephant hides, then the prehistoric hunting assumption could be questioned, which could result in new notions about the association between Clovis points and mammoth remains, such as that Clovis points were butchering tools. Frison's (1989, 783) tests showed that indeed "Clovis projectile points used with either atlatl and dart or thrusting spear will penetrate elephant hide and inflict lethal wounds on African elephants of all ages and both sexes."

Another study of tool effectiveness was Machin, Hosfield, and Mithen's (2007) examination of the relationship between symmetry and butchering effectiveness of Acheulean handaxes: "If a positive relationship exists [i.e., symmetry increases the effectiveness of a handaxe as a butchery tool], support can be given to those who argue that handaxes were primarily, or perhaps solely, subsistence tools. If no such relationship exists, then support will be given to those who argue that social, sexual, or aesthetic factors may have been important influences on handaxe morphology" (Machin, Hosfield, and Mithen 2007, 883). Effectiveness was measured in two ways. First, Machin and colleagues recorded the speed of the butchering event. Second, they assessed the "quality" of the event by asking their two test subjects, a professional game butcher and an archaeologist who studies the Paleolithic, to score the effectiveness of a used handaxe on several ordinal scales relating to different measures of quality and tool use. The plan-view and profile-view symmetry of each handaxe was quantified and then compared to the two effectiveness measures.

Machin, Hosfield, and Mithen's (2007) results were not as straightforward as Frison's (1989). There was only moderate support for the hypothesis that increasing plan-view symmetry increases the effectiveness of handaxes as butchering tools. Some tests were significant only for test butcher number 1, other tests were significant only for test butcher number 2, and many of the relationships between symmetry and butchery effectiveness, while significant, were weak. There was no support for the hypothesis that profile-view symmetry increased butchering effectiveness. Machin, Hosfield, and Mithen (2007, 892) concluded that their null hypothesis was better supported, namely that "factors other than functional considerations for animal butchery are playing a key role in the decisions by hominin stone knappers to impose high degrees of symmetry on some of their handaxes." However, they could not entirely rule out the butchering benefits of symmetry, which is interesting in and of itself.

We point out that as per the definition of our "replication experiment as test" category, neither Frison (1989) nor Machin, Hosfield, and Mithen (2007) directly compared their results to archaeological data, but important inferences regarding technology, technological evolution, and prehistoric behavior were made nonetheless. These inferences were not about tools from one Clovis site or one style of handaxe but instead were intended to be broader in scope and potentially applicable to interpretations of the Clovis period or the Lower Paleolithic period as a whole. That said, some questions are more specific and limited in applicability and thus require replication tests incorporating particular experimental variables. For example, Nigra and Arnold (2013) were interested in the production of beads from the shells of *Olivella biplicata* during the second millennium CE by Chumash hunter–gatherers on the California Channel Islands. They hypothesized that Chumash bead-production specialists chose Santa Cruz Island chert because "it demonstrated superior material properties for shell drilling" (Nigra and Arnold 2013, 3648). To test this hypothesis, Nigra and Arnold devised a series of experimental trials to compare locally available Santa Cruz Island chert against three alternative lithic materials (fused shale, obsidian, and Vandenberg/Monterey chert), all of which are available at prominent outcrops in southern California. They replicated the region's Middle-period-style Chumash flake drills, hafted the experimental drills in traditional fashion, and then tested them on *Olivella* shells in controlled laboratory conditions, examining two attributes: (1) the resilience of each raw material measured by the reduction in length of the drill after three minutes of drilling at three pounds of pressure and (2) the effectiveness of each raw material determined by whether a drill successfully perforated a bead blank. Analysis showed that Santa Cruz Island chert was both more resilient and effective than any of the other material types, lending strong inferential support for prehistoric preference for it.

Replication as model

Clarke (1972, 1) observed that "models are pieces of machinery that relate observations to theoretical ideas," although as Lycett and Chauhan (2010, 11–12) point out, the term "model" is frequently misused, and the purpose of a model is frequently misunderstood:

> Models are not by themselves statements about reality; rather they are formalized means of laying down explicit parameters in order that we can ask *how much does reality match this pattern?* Sometimes it will match the pattern with high degrees of fit; on other occasions, it will not match the data very well at all. Either way, we have made a manifest advance in our knowledge, being able to rule out or confirm the role of specific parameters and their strength of influence over a set of known variables. (emphasis in original)

Of the three types of models that Lycett and Chauhan (2010) describe, their "analogue model" is the most appropriate for our purposes here: "analogue models explicitly use information from better known or empirically documented situations (e.g. experiment or ethnography) to generate predictions. It is this sense of analogy between one set of empirical phenomena and another from which this subset of models takes its name." (p. 10)

Using stone-tool-replication experiments to help identify specific prehistoric reduction sequences and possible production behaviors has long been a staple of lithic analysis[4] (Akerman 2007; Aubry et al. 2008; Bradley and Sampson 1986; Clarkson, Shipton, and Weisler 2015b; Driscoll and García-Rojas 2014; Eren and Bradley 2009; Reti 2014; Schindler and Koch 2012; Shipton, Petraglia, and Paddayya 2009; Shott et al. 2007; Sollberger and Patternson 1976; Stafford 2003; Stout et al. 2014; Tryon, McBrearty, and Texier 2005; Wenban-Smith 1989). Although many of these analog models for production behaviors, both past and present, have relied on simple visual comparisons, increasingly sophisticated quantitative analyses (e.g. morphometrics) are providing an independent means for ensuring that comparisons between replicated and archaeological specimens are objective and robust. However, we cannot determine how challenging or difficult producing a stone-tool type or using a particular production technique would have been to a prehistoric knapper, who, unlike modern knappers, may have spent his or her life making and using stone tools, may have started learning at a much different (likely younger) age, and may have been surrounded by teachers or peers who had already learned the "trick" necessary to achieve production success (Eren et al. 2014a). Thus, caution and restraint should be exercised when it comes to specific proposals about a prehistoric person's stone-tool production learning trajectory, perception, specialization, or skill mastery.

The investigation of natural versus functional lithic fracture patterns, taphonomy, and use-wear studies has traditionally made use of "crash dummies"—replicates that are subjected to various processes such as butchering, projectile-shooting, and trampling and then compared to archaeological specimens (Andrefsky 2013; Bello, Parfitt, and Stringer 2009; Claud et al. 2015; Driscoll et al. 2015; Eren et al. 2010a, 2011a; Iovita et al. 2014; Jennings 2011; de Juana, Galán, and Domínguez-Rodrigo 2010; Key 2013; Lemorini et al. 2014; Lombard and Pargeter 2008; MacDonald 2014; Miller 2015; Pargeter and Bradfield 2012; Pevny 2012; Price 2012; Smallwood 2013; Tallavaara et al. 2010; Temple and Lee Sappington 2013; Weitzel et al. 2014a, 2014b). Heat treatment or damage has also been examined through stone-tool replication as model (Brown et al. 2009; Schmidt et al. 2013), although there are also examples of heat-treatment studies conducted through replication as test, in which researchers were more concerned with understanding the general process and parameters of the effect of heat on stone (Mercieca and Hiscock 2008; Schmidt et al. 2012, 2013).

One recent example of an experimental model is that of Wilkins et al. (2012), who hypothesized that diagnostic impact fractures on triangular stone flakes indicated that spear points were being produced at Kathu Pan 1, South Africa, 500,000 years ago. To test this hypothesis, experimentally replicated spearpoints made from the same banded ironstone as the archaeological specimens were thrust into animal carcasses with a calibrated crossbow "to simulate a thrusting spear and keep force constant" (Wilkins et al. 2012, 943). These experimental specimens subsequently served as a model against which the archaeological triangular flakes could be compared, which in turn were used to argue for support of the hypothesis. Although the results have been debated (McPherron et al. 2014; Rots and Plisson 2014; Wilkins et al. 2015), Wilkins et al.'s (2012) study illustrates the successful use of replication as model—that is, as a formalized means of establishing explicit parameters in order to compare archaeological data against modern, replicable data (Lycett and Chauhan 2010).

Wilkins et al. (2012) is one of several studies relating the identification of hunting function of stone artifacts to macrofractures. Over the past three decades at least 28 experimental projects on six continents have demonstrated, often independently, that a distinct subset of macrofractures, known as impact fractures or "diagnostic" impact fractures, form as a result of stone and bone tools being projected into animal carcasses (Table 1). This work has made model-hunting macrofactures one of the most investigated and replicated experimental stone-tool subjects. Most of these projects have further demonstrated that these fracture types are robust models for fractures that can be found on archaeological specimens (e.g. Barton & Bergman 1982; Fischer et al. 1984; Lombard and Pargeter 2008). Yet, identification of hunting macrofractures via experimental models is not without flaws. Notable issues include the use of a diverse range of nomenclature to refer to impact fractures; a lack of clear and accurate published images showing fracture initiations and terminations; and a frequent lack of macrofracture quantification. Most recently, research teams working independently of one another in South Africa (e.g. Pargeter 2011; Pargeter and Bradfield 2012) and Japan (e.g. Sano 2009) have reached convergent conclusions regarding the frequencies (c. < 5%) of impact fractures likely to form under non-hunting conditions (Table 1). These results have made the publishing of impact fracture frequencies a requisite.

Experimentally derived replication models are not always in the form of replicated specimens being compared to archaeological specimens. For example, Eren and Andrews (2013) were interested in whether Clovis foragers in the North American Great Lakes region transported large biface cores and produced stone flakes on the go or instead knapped their flakes at a stone source before setting off on treks. Flake thickness is minimally affected by retouch and resharpening (Patten 2005; Shott and Weedman 2007; Surovell 2009), so Eren and Andrews reasoned that if they could understand how flake thickness is patterned over the course of biface-core reduction, it would be possible to construct predictions (a model) to infer from a sample of sites whether or not bifaces were being transported as mobile cores.

To understand the patterning of blank thickness, they replicated Clovis biface cores (Figure 2) and plotted the sequence of flake removal against flake thickness. They found two trends: (1) a significant negative linear relationship between the sequence of flake removal and flake thickness (Figure 3a) and (2) a significant negative linear relationship between sequence of flake removal and variation in flake thickness (Figure 3b). They then predicted that if foragers carried their cores with them, core reduction should progressively advance to later stages of knapping the farther they traveled away from a stone source. Thus in a group of sites possessing

TABLE 1

OVERVIEW OF EXPERIMENTAL STUDIES INVESTIGATING IMPACT FRACTURE FORMATION ON STONE TOOLS USED AS HUNTING WEAPONS AND DURING TAPHONOMIC AND TOOL PRODUCTION PROCESSES.

Author	Year	Lab/actualistic	Test type
Barton and Bergman	1982	Actualistic	Arrows
Huckell	1982	Actualistic	Spears
Moss and Newcomer	1982	Actualistic	Arrows
Bergman and Newcomer	1983	Actualistic	Arrows
Fischer et al.	1984	Actualistic	Spears, arrows
Odell and Cowan	1986	Actualistic	Spears, arrows
Albarello	1986	Actualistic	Arrows
Titmus and Woods	1986	Actualistic	Darts
Geneste and Plisson	1990	Actualistic	Spears, arrows
Caspar and De Bie	1996	Actualistic	Arrows
Soriano	1998	Actualistic	Arrows
Plisson and Beyries	1998	Actualistic	Spears
Kelterborn	1999	Actualistic	Arrows
Crombé et al.	2001	Actualistic	Arrows
Shea et al.	2002	Actualistic	Spears
Lombard et al.	2004	Actualistic	Spears
O'Farrell	2004	Actualistic	Spears
Lombard and Pargeter	2008	Lab/actualistic	Spears
Yaroshevich et al.	2010	Actualistic	Arrows
Flegenheimer, Martínez, and Colombo	2010	Actualistic	Spears
Brindley	2011	Actualistic	Spears
Petillon et al.	2011	Actualistic	Spears
Wilkins et al.	2012	Actualistic	Spears
Iovita et al.	2014	Lab	Spears
Weitzel et al.	2014	Actualistic	Spears, darts
Sano and Oba	2014	Lab/actualistic	Arrows, spears
Taphonomic/production related impact fracture experiments			
Fisher et al.	1984	Actualistic	Knapping, trampling
O'Farrell	2004	Actualistic	Knapping, trampling
Sano	2009	Actualistic	Trampling
Weitzel	2010	Actualistic	Knapping, trampling
Pargeter	2011	Actualistic	Trampling
Pargeter and Bradfield	2012	Actualistic	Trampling, knapping
Pargeter	2013	Actualistic	Trampling, stone rolling, dropping

Note that this table covers only studies with published impact fracture descriptions or frequencies.

FIGURE 2 A replica Clovis biface core and flakes knapped by M. I. Eren. The top row of flakes are from earlier in the reduction sequence and, as a group, are thicker and more variable in their thicknesses. The bottom row of flakes are from later in the reduction sequence and, as a group, are thinner and less variable in their thicknesses. These trends in the experimental replication were used to construct a mathematical model that could be directly compared against data from the archaeological record (see text and Figure 3).

various site-to-source distances, there should be negative relationships between distance and unifacial-tool (flake) thickness (Figure 3c) and between distance and variation (standard deviation) in thickness (Figure 3d). When archaeological data were compared against the analog model (Figure 3e, f), they showed the opposite pattern, and Eren and Andrews (2013) rejected the model and concluded that Clovis foragers in the North American Great Lakes did not carry biface cores with them.

Replication as method validation

In this category of replication, experimentally knapped specimens are used as control groups to assess quantitative methods that will ultimately be used on archaeological specimens. After some early work (Kuhn 1990), this use of replication has increased substantially over the last 15 years. One avenue of inquiry—creating

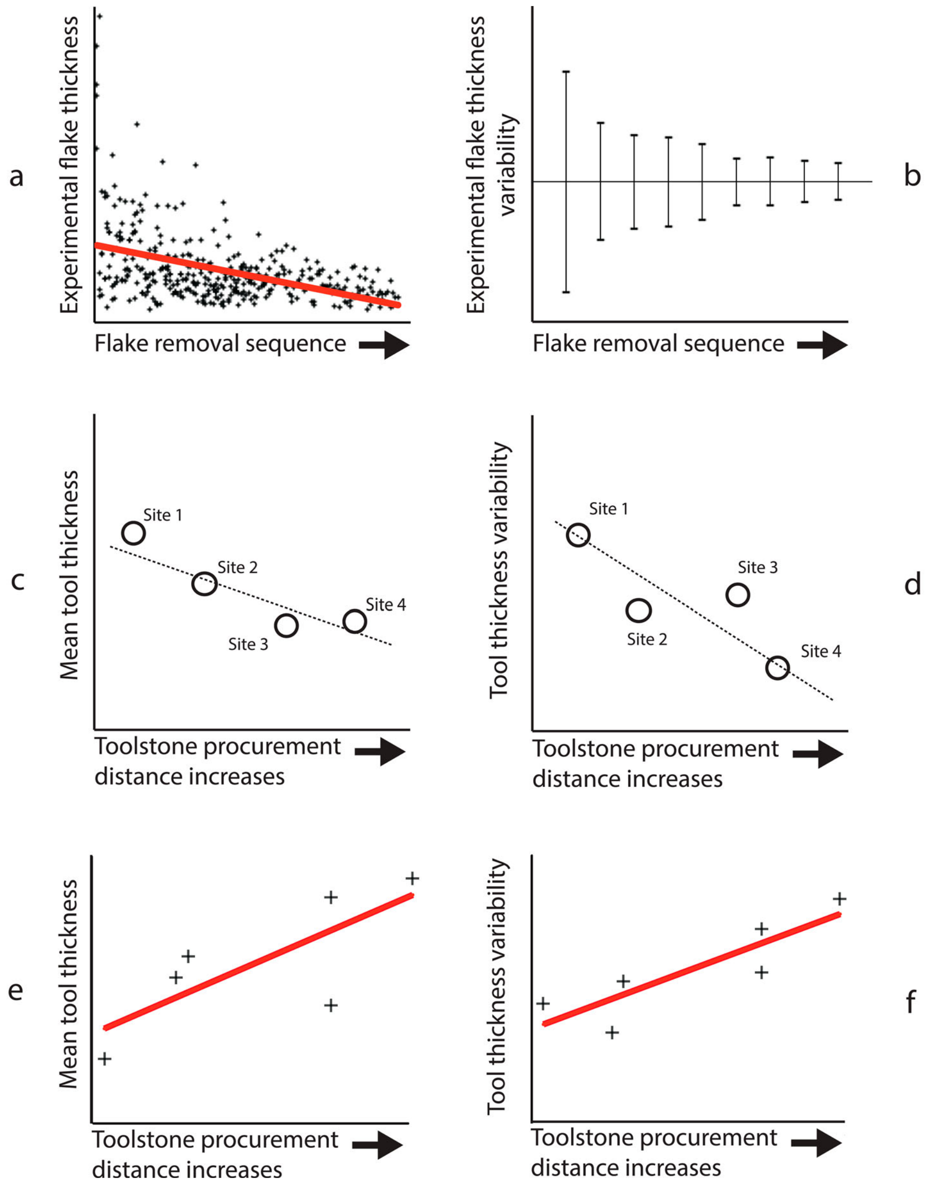

FIGURE 3 From experimental replication of Clovis biface cores (see Figure 2), Eren and Andrews (2013) uncovered two simple, but important, trends (a, b). These two trends allowed them to create a model (c, d) to test against the archaeological record (e, f) whether Clovis Paleoindians carried their biface cores.

methods for estimating the amount or effects of mass removed from a flake or tool by retouch or resharpening—has received extensive coverage in the literature (Andrefsky 2006; Bradbury, Carr, and Randall Cooper 2009; Braun et al. 2010; Clarkson 2002; Davis and Shea 1998; Eren et al. 2005; Eren and Prendergast

2008; Eren and Sampson 2009; Hiscock and Clarkson 2005; Horowitz and McCall 2013; Patten 2005; Shott et al. 2000, 2007; Marwick 2008; Morales, Lorenzo, and Vergès 2015; Wilson and Andrefsky 2008), but with some exceptions, it has not been intensive. Rather, researchers have focused on using an experiment for the purpose of proposing new or newly revamped methods rather than for thoroughly vetting methods and validating that they actually work for their asserted purpose (Shott et al. 2007:205–206).

For other methodological topics, such as those dealing with cortex (Dibble et al. 2005), edge length (Mackay 2008), core reduction, and flake-scar density (Clarkson 2013; Clarkson, Shipton, and Weisler 2015c), even fewer experimental validations have been conducted. To be clear, the responsibility of method validation is discipline wide and falls to anyone interested in potentially applying a method. Further, not all methods require a great deal of experimentation before researchers can confidently use them. However, a major problem arises when either the original authors of a method, or other researchers who use said method, are not patient enough to conduct, or wait for, the necessary sets of experiments to robustly validate a method that may very well need scrutiny. This may be especially the case where a wide array of variables present in the archaeological record may have a direct bearing on the sensitivity and accuracy of the method. Some methods may require continuous and alternating experimentation and application in a cyclical fashion, especially when applied to new or different kinds of archaeological data (Andrefsky 2007; Bradbury and Carr 2009, 2795).

Experimentally replicated stone specimens can serve as useful elements for establishing or better understanding optimal or satisfactory protocols for methods such as lithic refitting. For example, Laughlin and Kelly (2010) conducted an experiment that examined the effects of experience/aptitude, type of reduction (biface versus core), and flake-size cut-offs on different rates of success in refitting. Using production debitage from replicated bifaces and cores of Wyoming Green River chert, Laughlin and Kelly recruited 13 people to refit as many flakes as possible within two hours. The results confirmed that all three variables played a significant role in refitting success, which in turn allowed Laughlin and Kelly to make several recommendations about productive ways of conducting a refitting study. Perhaps even more valuable was the presentation of cumulative refit curves that allow refitters to figure out when they have reached a point of diminishing returns in terms of number of successful refits from an archaeological assemblage.

Future challenges

Archaeologists who use replication experiments face several challenges (Kelly 1994). Researchers must always be wary of the "flintknapper's fundamental conceit" (Thomas 1986). There is an inherent danger that archaeological flintknappers might exploit their intuitive knowledge of stone-tool replication as an authoritative trump card to overrule colleagues who are not flintknappers — or as a tactic for influencing the public into believing that because they understand how to make stone tools, they automatically understand prehistoric forager behavior, evolution, adaptation, dispersal, and culture as well. Alternatively, and perhaps in reaction to this latter behavior, there are archaeologists who dismiss the usefulness of any stone-tool replication experiments — this second attitude is unreasonable as well. While opposed, both of these viewpoints stem from a poor articulation with the principle of uniformitarianism. The first "intuitive" view exaggerates the principle of uniformitarianism to such an extent that a scientific framework no longer becomes necessary to test hypotheses, the knapper simply "knows" the past because he or she is "reproducing it." The second "reactionary" view ignores the fact that stone breaks the same way today as it did in the past and possesses the same physical properties as it did in the past (sharp cutting edge, durability, morphology, and so on), readily facilitating some level of uniformitarian link that is exploitable scientifically. That is, if one accepts that rocks in the past fractured similarly to rocks in the present, then it should go without saying that particular hypotheses and predictions about stone-tool efficiency, morphology, function, and other topics reviewed above, are of course validly examined via stone-tool replication experiments conducted within a scientific framework of test, model, or method validation.

Another challenge faced by archaeologists who use replication experiments is that some experimental protocols, which are standard in other disciplines, have yet to be widely applied in tool-replication experiments, much less scrutinized from the standpoint of when to use them. Take, for example, blind testing, which is not always necessary for hypothesis testing. Whether it is applied or not depends on the question being asked, and in some cases a question might benefit from tests employing both nonblind and blind trials. In nonblind trials, the potential for knapper bias, unconscious though it may be, is always present. Because of this, nonblind replication experiments should be viewed more along the lines of experimental computer or mathematical simulations, in which a programmer chooses which variables to include. To be sure, knapper (and programmer) bias in nonblind experiments can be minimized, curbed, or identified through the use of explicit, replicable instructions, such as instructing the knapper to stay within particular parameters or to copy a standard model (Eren et al. 2014b) or via comparisons of replicated and archaeological specimens (Eren and Lycett 2012; Shott 2002). That said, the use of blind testing is an important step forward, even if that means in some cases and for some questions expert knappers can no longer participate in the experiments they themselves designed (but see Nami 2010).

Replication experiments should always attempt to reach statistically valid sample sizes. For example, although Eren, Greenspan, and Sampson's (2008) experimental study of blade-reduction efficiency versus discoidal-core-reduction efficiency measured the cutting-edge length of thousands of specimens, the sample of core reductions—the topic of investigation—was small: the actual statistical comparisons encompassed only seven blade cores and seven discoidal cores. Of course, there are always practical considerations, and as anyone who has conducted a cutting-edge-efficiency experiment can attest, time is a major constraint. Jennings, Pevny, and Dickens' (2010) use of combined experimental datasets is one possible and productive way forward. Collaboration among researchers is another way to achieve valid sample sizes. Some stone-tool experiments require large samples not in terms of specimens but of participants. We touch on this issue below.

As we have discussed, inevitable tradeoffs involved in experimental design ensure that no single experiment will strike a "perfect" balance between realism and control, especially given inevitable financial and practical constraints typically facing an experimenter. When blind testing or large sample sizes cannot be immediately achieved, independent re-testing may support an analysis. Thus, Clarkson's (2010) independent experimental confirmation of Eren, Greenspan, and Sampson's (2008) results provides the latter with empirical support despite the small sample sizes and lack of blind testing. However, even when blind testing is conducted or large sample sizes are used, independent testing and re-testing is desirable. We are unsure as to why archaeologists are generally slow or reluctant to conduct independent experimental re-tests, but there may be concerns over spending time on an endeavor that is erroneously believed to not hold much prestige, or yield much "credit," in the eyes of the archaeological community, universities, or the academy in general. Perhaps there is an erroneous perception that one experiment has settled an issue. We also need more experiments that vary the experimental variables and parameters of published experiments. In terms of replication experiments for method validation, we need to move away from the widespread belief that one or two experiments validates a method, especially when experiments are not conducted blind (e.g. Rots and Plisson 2014, 158).

Another challenge that faces stone-tool replication is the permanent curation of replicated data sets. Whereas many current experiments are employing large numbers of specimens (Figure 4), and space always seems to be in short supply, archaeologists should nonetheless consider curating experimental materials. This will allow other researchers not only to examine the work that has been done but also to ask and answer new questions without having to create the data sets themselves. For these new questions, the curated experimentally replicated specimens might in some cases even act as blind tests, given that the person who generated the replicas might be unaware of the new study's goals (Gurtov, Buchanan, and Eren 2015). In cases where physical curation is impractical or impossible, 3D scanning and printing may allow experimental specimens to be curated digitally.

The challenges described above may stem from the fact that, despite the great strides made in recent years, experimental stone-tool replication is still in a state of scientific immaturity relative to other experimental sciences (e.g. Bradbury and Carr 2009, 2795). But these challenges are also in part due to more immediate, practical problems. It may be difficult for experimenters to get funding to pay for large numbers of participants in blind-trials, much less large numbers of participants who also happen to be highly skilled knappers, which is necessary if the topic of investigation is something like Preferential (lineal) Levallois reduction, Clovis fluted projectile-points, Danish Neolithic daggers, or Egyptian Gerzean knives. If they can find the time away from teaching, administrative duties, conducting research, and publishing, professional archaeologists and graduate students may very well be happy to participate in their colleagues' experiments for free, but unfortunately there is a current dearth of professional archaeologists and archaeology students who are also highly skilled flintknappers—becoming a skilled knapper requires a tremendous dedication and financial investment. The hobby knapping community might play an important role in this regard as potential test subjects, but we then

FIGURE 4 The dataset from Eren et al.'s (2014b) study "The role of raw material differences in stone-tool shape variation: an experimental assessment." While the initial published study focused on the 105 finished tool replicas (some of which are pictured to the right), all production flakes for each tool were weighed, numbered, carefully boxed, and saved for future study.

again face the issue of funding—paying participants in blind trials for their time and effort, as well as traveling to them or bringing them to our labs. And this of course assumes one lives in the United States; other countries do not necessarily have thriving communities of skilled hobby knappers.

Conclusion

Because of ever-increasing use of structured research design, hypothesis testing, quantitative methods, and inferential statistics, stone artifact replication is contributing to our knowledge of prehistoric behavior at an unprecedented rate. Still, those of us involved in replication experiments have our work cut out for us in terms of catching up with more mature experimental sciences such as biology, physics, or psychology. We say this because we see a tendency in archaeology to ignore issues such as external and internal validity and to substitute intuition for theoretically based hypotheses that clearly delineate independent and dependent variables. Here we have emphasized that stone-tool replication can be used in three productive ways: (1) to test a question, hypothesis, or assumption about certain parameters of stone-tool technology; (2) as a model, in which information from empirically documented situations, such as an experiment, is used to generate predictions; and (3) as a means of validating methods—for example, using experimentally knapped tools to assess quantitative methods that will be used on archaeological specimens. We hope our modest effort will help sort out some of the epistemological issues surrounding the use of replication experiments and spur the growth of hypothesis-driven studies.

Acknowledgments

We thank Matthew Boulanger, Philip Carr, Kathryn Kamp, Alastair Key, Kerstin Schillinger, John Whittaker, and three anonymous reviewers for reading over early drafts of this manuscript. J.P is supported by the Leakey Foundation Mosher Baldwin Fellowship and the National Science Foundation Doctoral Dissertation Improvement (Grant ID: 1542310). S.J.L is supported by the Research Foundation for the State University of New York. During the period this article was written M.I.E. was supported by a University of Missouri postdoctoral fellowship. M.I.E. dedicates this work to P. J. C.-E.

Notes

1 It should be noted that the terms internal and external validity were originally designed to be applied solely to experiments, and Lycett and Eren's (2013; see also: Roe and Just 2009; Mesoudi 2011) application of them to discuss alternative *research methods* inevitably involves some slight manipulation of these terms as they were originally conceived in the case of (solely) experiments. Nevertheless, use of this terminology highlights the contrasts in alternative strengths and weakness of these different research methods, which is useful in an inevitably historical science such as archaeology where the past cannot be "replicated."

2 The reader may be interested to note that some initial insights of individual flake fracture were first identified by Wilmsen (1970:67) over 40 years ago. Based on his analysis of North American Paleoindian flake tools, he inferred from his archaeological data "that striking platform architecture is of fundamental importance in predetermining at least some flake form characteristics. Overall specimen size, although probably related in part to raw material size, is also directly related to platform size. Platform thickness is apparently a strong determinant of specimen thickness and width and, to a lesser extent, of specimen length." He later writes: "While it is probably not true that a Paleo-Indian knapper could direct every single flake to a specific size and shape, it appears to be certain that s/he could regulate any series of flakes to meet intended dimensional and formal tolerances. S/he apparently did this by varying the distance from the edge of a core at which he applied detaching force…" He also investigates platform width and flake angle (interior platform angle), and speculates about force of blow.

3 In this regard, tool-making "exploration" by avocational or hobby flintknappers may at times be valuable to archaeological inquiry. E-forums for knapping enthusiasts, like the paleoplanet prehistoric skills forum (http://paleoplanet69529.yuku.com/), or YouTube videos (Eren et al. 2010b; Shea 2015), reveal the use of many tools and processes that have not received a lot of formal study archaeologically, including punch work and use of wooden billets. Discussions with avocational knappers could potentially indicate where further formal investigation might be fruitful and may help identify new variables, attributes, or behaviors to be investigated archaeologically, or in controlled replication experiments via human replication or machine/device flaking.

4 Many experimentally replicated analogue models for lithic reduction sequences and production behaviors are never actually compared – via eyeballs or otherwise – to the archaeological data they are purportedly interested in better understanding or explaining. These "orphan models" could potentially serve as great source of graduate student research and publication by quantitatively comparing whether what was replicated experimentally actually matches archaeological data. Similarly, many taphonomy-focused experiments using replicated stone tool test specimens

serve merely as "cautionary tales" or "discoveries of note" in that the data and patterns generated by the experiment *could* be used as a potential model against which archaeological data are compared but are not, at least not immediately. For example, in their experimental comparison of stone-flake versus bamboo cutmark morphology, West and Louys (2007) suggested that they found differences that could possibly be identified archaeologically. Yet, to our knowledge, no one has compared West and Louys' (2007) experimentally derived model to archaeological data.

References

Ahler, Stanley. 1989a. "Experimental Knapping with KRF and Midcontinent Cherts: Overview and Applications." In *Experiments in Lithic Technology*, edited by Daniel S. Amick, and Raymond P. Mauldin, 67–99. Oxford: BAR International Series 528.

Akerman, Kim. 2007. "To Make a Point: Ethnographic Reality and the Ethnographic and Experimental Replication of Australian Macroblades Known as Leilira." *Australian Archaeology* 64: 23–34.

Albarello, Bruno. 1986. "Sur l'usage des microlithes comme armatures de projectiles." *Revue Archéologique du Centre de la France* 25: 127–143.

Andrefsky, William, Jr. 2006. "Experimental and Archaeological Verification of an Index of Retouch for Hafted Bifaces." *American Antiquity* 71:743–757.

———. 2007. "The Application and Misapplication of Mass Analysis in Lithic Debitage Studies." *Journal of Archaeological Science* 34:392–402.

———. 2013. "Fingerprinting Flake Production and Damage Processes: Toward Identifying Human Artifact Characteristics." In *Paleoamerican Odyssey*, edited by Kelly E. Graf, Caroline V. Ketron, and Michael R. Waters, 415–428. College Station: Texas A&M University.

Archer, Will, and David R. Braun. 2010. "Variability in Bifacial Technology at Elandsfontein, Western Cape, South Africa: A Geometric Morphometric Approach." *Journal of Archaeological Science* 37: 201–209.

Aubry, Thierry, Bruce Bradley, Miguel Almeida, Bertrand Walter, Maria J. Neves, Jacques Pelegrin, Michel Lenoir, and Marc Tiffagom. 2008. "Solutrean Laurel Leaf Production at Maîtreaux: An Experimental Approach Guided by Techno-Economic Analysis." *World Archaeology* 40: 48–66.

Barham, Lawrence. 1987. "The Bipolar Technique in Southern Africa: A Replication Experiment." *South African Archaeological Bulletin* 42: 45–50.

Bar-Yosef, Ofer, Metin I. Eren, Jiarong Yuan, David J. Cohen, and Yiyuan Li. 2012. "Were Bamboo Tools Made in Prehistoric Southeast Asia? An Experimental View from South China." *Quaternary International* 269: 9–21.

Barton, Nick, and Christopher A. Bergman. 1982. "Hunters at Hengistbury: Some Evidence from Experimental Archaeology." *World Archaeology* 14: 237–248.

Bergman, A Christopher, and Mark H. Newcomer. 1983. "Flint Arrowhead Breakage: Examples from Ksar Akil, Lebanon." *Journal of Field Archaeology* 10: 238–243.

Brindley, Jared. 2011. Beyond a Suggestive Morphology: A Study of Point Use in Wardaman Country, Northern Territory, Australia., Bachelor of Arts Honours (Archaeology) in the School of Social Science, University of Queensland.

Bello, Silvia M., Simon A. Parfitt, and Chris Stringer. 2009. "Quantitative Micromorphological Analyses of Cut Marks Produced by Ancient and Modern Handaxes." *Journal of Archaeological Science* 36: 1869–1880.

Bradbury, Andrew P., and Philip J. Carr. 2009. "Hits and Misses When Throwing Stones at Mass Analysis." *Journal of Archaeological Science* 36: 2788–2796.

Bradbury, Andrew P., Philip J. Carr, and D. Randall Cooper. 2009. "Raw Material and Retouched Flakes." In *Lithic Technology: Measures of Production, Use, and Curation*, edited by William Andrefsky, Jr., 233–254. Cambridge: University of Cambridge Press.

Bradley, Bruce, and C. Garth Sampson. 1986. "Analysis by Replication of Two Acheulian Artefact Assemblages." In *Stone Age Prehistory: Studies in Memory of Charles McBurney*, edited by G.N. Bailey, and P. Callow, 29–45. Cambridge: Cambridge University Press.

Braun, David R., Briana L. Pobiner, and Jessica C. Thompson. 2008. "An Experimental Investigation of Cut Mark Production and Stone Tool Attrition." *Journal of Archaeological Science* 35: 1216–1223.

Braun, David R., Thomas Plummer, Joseph V. Ferraro, Peter Ditchfield, and Laura C. Bishop. 2009. "Raw Material Quality and Oldowan Hominin Toolstone Preferences: Evidence from Kanjera South, Kenya." *Journal of Archaeological Science* 36: 1605–1614.

Braun, David R., Michael J. Rogers, John W. Harris, and Steven J. Walker. 2010. "Quantifying Variation in Landscape-Scale Behaviors: The Oldowan of Koobi Fora." In *New Perspectives on Old Stones*, edited by Stephen J. Lycett, and Parth Chauhan, 167–182. Dordrecht: Springer.

Brown, Kyle S., Curtis W. Marean, Andy I.R. Herries, Zenobia Jacobs, Chantal Tribolo, David Braun, David L. Roberts, Michael C. Meyer, and Jocelyn Bernatchez. 2009. "Fire as an Engineering Tool of Early Modern Humans." *Science* 325: 859–862.

Cahen, Daniel. 1978. "Comment on A History of Flint-Knapping Experimentation, 1838–1976." *Current Anthropology* 19: 360.

Carr, Philip J., and Andrew P. Bradbury. 2010. "Flake Debris and Flintknapping Experimentation." In *Designing Experimental Research in Archaeology: Examining Technology Through Production and Use*, edited by Jeffrey R. Ferguson, 71–91. Boulder: University Press of Colorado.

Caspar, Jean-Paul, and Mark De Bie. 1996. "Preparing for the Hunt in the Late Paleolithic Camp at Rekem, Belgium." *Journal of Field Archaeology* 23: 437–460.

Chazan, Michael. 2013. "Butchering with Small Tools: The Implications of the Evron Quarry Assemblage for the Behaviour of *Homo Erectus*." *Antiquity* 87: 350–367.

Cheshier, Joseph, and Robert L. Kelly. 2006. "Projectile Point Shape and Durability: The Effect of Thickness: Length." *American Antiquity* 71: 353–363.

Clarke, David L. 1968. *Analytical Archaeology*. London: Methuen.

Clarke, David L. (Editor) 1972. *Models in Archaeology*. London: Methuen.

Clarkson, Chris. 2002. "An Index of Invasiveness for the Measurement of Unifacial and Bifacial Retouch: A Theoretical, Experimental, and Archaeological Verification." *Journal of Archaeological Science* 29: 65–75.

——. 2010. "Regional Diversity Within the Core Technology of the Howiesons Poort Techno-Complex." In *New Perspectives on Old Stones*, edited by Stephen J. Lycett, and Parth Chauhan, 43–60. Dordrecht: Springer.

——. 2013. "Measuring Core Reduction using 3D Flake Scar Density: A Test Case of Changing Core Reduction at Klasies River Mouth, South Africa." *Journal of Archaeological Science* 40: 4348–4357.

Clarkson, Chris, Michael Haslam, and Clair Harris. 2015a. "When to Retouch, Haft, or Discard? Modeling Optimal Use/Maintenance Schedules in Lithic Tool Use." In *Lithic Technological Systems and Evolutionary Theory*, edited by Nathan Goodale, and William Andrefsky, Jr., 117–138. Cambridge: Cambridge University Press.

Clarkson, Chris, Ceri Shipton, and Marshall Weisler. 2015b. "Front, Back, and Sides: Experimental Replication and Archaeological Analysis of Hawaiian Adzes and Associated Debitage." *Archaeology in Oceania* 50: 71–84.

Clarkson, Chris, Ceri Shipton, and Marshall Weisler. 2015c. "Determining the Reduction Sequence of Hawaiian Quadrangular Adzes using 3D Approaches: A Case Study from Moloka'i." *Journal of Archaeological Science* 49: 361–371.

Claud, Émilie, Marianne Deschamps, David Colonge, Vincent Mourre, and Céline Thiébaut. 2015. "Experimental and Functional Analysis of Late Middle Paleolithic Flake Cleavers from Southwestern Europe (France and Spain)." *Journal of Archaeological Science* 62: 105–127.

Collins, Sophie. 2008. "Experimental Investigations into Edge Performance and its Implications for Stone Artefact Reduction Modelling." *Journal of Archaeological Science* 35: 2164–2170.

Crombé, Philippe, Yves Perdaen, Joris Sergant, and Jean-Paul Caspar. 2001. "Wear Analysis on Early Mesolithic Microliths from the Verrebroek Site, East Flanders, Belgium." *Journal of Field Archaeology* 28: 253–269.

Davis, Zachary J., and John J. Shea. 1998. "Quantifying Lithic Curation: An Experimental Test of Dibble and Pelcin's Original Flake-Tool Mass Predictor." *Journal of Archaeological Science* 25: 603–610.

Dibble, Harold L. 1997. "Platform Variability and Flake Morphology: A Comparison of Experimental and Archaeological Data and Implications for Interpreting Prehistoric Lithic Technological Strategies." *Lithic Technology* 22: 150–170.

——. 1998. "Comment on "Quantifying Lithic Curation: An Experimental Test of Dibble and Pelcin's Original Flake-Tool Mass Predictor," by Zachary J. Davis and John J. Shea." *Journal of Archaeological Science* 25: 611–613.

Dibble, Harold L., and Andrew Pelcin. 1995. "The Effect of Hammer Mass and Velocity on Flake Mass." *Journal of Archaeological Science* 22: 429–439.

Dibble, Harold L., and Zeljko Rezek. 2009. "Introducing a New Experimental Design for Controlled Studies of Flake Formation: Results for Exterior Platform Angle, Platform Depth, Angle of Blow, Velocity, and Force." *Journal of Archaeological Science* 36: 1945–1954.

Dibble, Harold L., and John C. Whittaker. 1981. "New Experimental Evidence on the Relation Between Percussion Flaking and Flake Variation." *Journal of Archaeological Science* 8: 283–296.

Dibble, Harold L., Utsav A. Schurmans, Radu P. Iovita, and Michael V. McLaughlin. 2005. "The Measurement and Interpretation of Cortex in Lithic Assemblages." *American Antiquity* 70: 545–560.

Diez-Martín, Fernando, Policarpo Sánchez-Yustos, Manuel Domínguez-Rodrigo, and Mary E. Prendergast. 2011. "An Experimental Study of Bipolar and Freehand Knapping of Naibor Soit Quartz from Olduvai Gorge (Tanzania)." *American Antiquity* 76: 690–708.

Diez-Martin, Fernando, and Metin I. Eren. 2012. "The Early Acheulean in Africa: Past Paradigms, Current Ideas, and Future Directions." In *Stone Tools and Fossil Bones: Debates in the Archaeology of Human Origins*, edited by Manuel Domínguez-Rodrigo, 310–358. Cambridge: Cambridge University Press.

Dincauze, Dena. 1978. "Comment on A History of Flint-Knapping Experimentation, 1838–1976." *Current Anthropology* 19: 360.

Driscoll, Killian. 2011. "Vein Quartz in Lithic Traditions: An Analysis Based on Experimental Archaeology." *Journal of Archaeological Science* 38: 734–745.

Driscoll, Killian, and Maite García-Rojas. 2014. "Their Lips Are Sealed: Identifying Hard Stone, Soft Stone, and Antler Hammer Direct Percussion in Palaeolithic Prismatic Blade Production." *Journal of Archaeological Science* 47: 134–141.

Driscoll, Killian, Jonas Alcaina, Natàlia Égüez, Xavier Mangado, Josep-Maria Fullola, and José-Miguel Tejero. 2015. "Trampled Under Foot: A Quartz and Chert Human Trampling Experiment at the Cova del Parco Rockshelter, Spain." *Quaternary International*: In Press.

Duke, Hilary, and Justin Pargeter. 2015. "Weaving Simple Solutions to Complex Problems: An Experimental Study of Skill in Bipolar Cobble-Splitting." In *Exploring Variability in Bipolar Technology*, edited by Justin Pargeter, Hilary Duke. Special Publication, *Lithic Technology* 40: 349–365.

Dunnell, Robert. 1971. *Systematics in Prehistory.* New York: The Free Press.

Eren, Metin I., and Brian N. Andrews. 2013. "Were Bifaces Used as Mobile Cores by Clovis Foragers in the North American Lower Great Lakes Region? An Archaeological Test of Experimentally-Derived Quantitative Predictions." *American Antiquity* 78: 166–180.

Eren, Metin I., and Bruce Bradley. 2009. "Experimental Evaluation of the Levallois "Core Shape Maintenance" Hypothesis." *Lithic Technology* 34: 119–125.

Eren, Metin I., and Mary E. Prendergast. 2008. "Comparing and Synthesizing Lithic Reduction Indices." In *Lithic Technology: Measures of Production, Use, and Curation*, edited by William Andrefsky, Jr., 49–85. Cambridge: University of Cambridge Press.

Eren, Metin I., and C. Garth Sampson. 2009. "Kuhn's Geometric Index of Unifacial Stone Tool Reduction (GIUR): Does it Measure Missing Flake Mass?" *Journal of Archaeological Science* 36: 1243–1247.

Eren, Metin I., Manuel Domínguez-Rodrigo, Ian Le, Daniel S. Adler, Steven L. Kuhn, and Ofer Bar-Yosef. 2005. "Defining and Measuring Reduction in Unifacial Stone Tools." *Journal of Archaeological Science* 32: 1190–1201.

Eren, Metin I., Aaron Greenspan, and C. Garth Sampson. 2008. "Are Upper Paleolithic Blade Cores more Productive than Middle Paleolithic Discoidal Cores? A Replication Experiment." *Journal of Human Evolution* 55: 952–961.

Eren, Metin I., Adam Durant, Christina Neudorf, Michael Haslam, Ceri Shipton, Janardhana Bora, Ravi Korisettar, and Michael Petraglia. 2010a. "Experimental Examination of Animal Trampling Effects on Artifact Movement in Dry and Water Saturated Substrates: A Test Case from South Indian." *Journal of Archaeological Science* 37: 3010–3021.

Eren, Metin I., Mark Kollecker, Chris Clarkson, and Bruce Bradley. 2010b. "Dynamic Approaches to Teaching Lithic Technology: Using YouTube in the Classroom." *Ethnoarchaeology* 2: 223–234.

Eren, Metin I., Andrew Boehm, Brooke Morgan, Richard Anderson, and Brian N. Andrews. 2011a. "Flaked Stone Taphonomy: A Controlled Experimental Study of the Effects of Sediment Consolidation on Flake Edge Morphology." *Journal of Taphonomy* 9: 201–217.

Eren, Metin I., Bruce Bradley, and C. Garth Sampson. 2011b. "Middle Paleolithic Skill-level and the Individual Knapper: An Experiment." *American Antiquity* 76: 229–251.

Eren, Metin I., Stephen J. Lycett, Christopher I. Roos, and C. Garth Sampson. 2011c. "Toolstone Constraints on Knapping Skill: Levallois Reduction with Two Different Raw Materials." *Journal of Archaeological Science* 38: 2731–2739.

Eren, Metin I., and Stephen J. Lycett. 2012. "Why Levallois? A Morphometric Comparison of Experimental "Preferential" Levallois Flakes Versus Debitage Flakes." *PLoS ONE* 7 (1): e29273.

Eren, Metin I., Robert J. Patten, Michael J. O'Brien, and David J. Meltzer. 2013. "Refuting the Technological Cornerstone of the Ice-Age Atlantic Crossing Hypothesis." *Journal of Archaeological Science* 40: 2934–2941.

——. 2014a. "More on the Rumor of "Intentional Overshot Flaking" and the Purported Ice-Age Atlantic crossing." *Lithic Technology* 39: 55–63.

Eren, Metin I., Christopher I. Roos, Brett Story, Noreen von Cramon-Taubadel, and Stephen J. Lycett. 2014b. "The Role of Raw Material Differences in Stone Tool Shape Variation: An Experimental Assessment." *Journal of Archaeological Science* 49: 472–487.

Faisal, Aldo, Dietrich Stout, Jan Apel, and Bruce Bradley. 2010. "The Manipulative Complexity of Lower Paleolithic Stone Toolmaking." *PLoS ONE* 5: e13718.

Fischer, Anders, Peter Vemming Hansen, and Peter Rasmussen. 1984. "Macro and Microwear Traces on Lithic Projectile Points: Experimental Results and Prehistoric Examples." *Journal of Danish Archaeology* 3: 19–46.

Flegenheimer, Nora, Jorge G. Martínez, and Mariano Colombo. 2010. "Una experiencia de lanzamiento de puntas cola de pescado." In *Mamül Mapu: Pasado y Presente desde la Arqueología Pampeana. Editorial Libros del Espinillo*, edited by M. Berón, L. Luna, M. Bonomo, C. Montalvo, C. Aranda, and M.C. Aizpitarte, 215–236. Buenos Aires.

Flenniken, J. Jeffrey. 1984. "The Past, Present, and Future of Flintknapping: An Anthropological Perspective." *Annual Review of Anthropology* 13: 187–203.

Flenniken, J. Jeffrey, and Anan W. Raymond. 1986. "Morphological Projectile Point Typology: Replication Experimentation and Technological Analysis." *American Antiquity* 51: 603–614.

Friis-Hansen, Jan. 1990. "Mesolithic Cutting Arrows: Functional Analysis of Arrows Used in the Hunting of Large Game." *Antiquity* 64: 494–504.

Frison, George C. 1989. "Experimental Use of Clovis Weaponry and Tools on African Elephants." *American Antiquity* 54: 766–784.

Galán, A.B., and M. Domínguez-Rodrigo. 2014. "Testing the Efficiency of Simple Flakes, Retouched Flakes and Small Handaxes During Butchery." *Archaeometry* 56: 1054–1074.

Geneste, Jean-Michel, and Hughes Plisson. 1990. "Technologie fonctionelle des pointes a cran Solutréennes: l'apport des nouvelles données de la grotte de Combe Saunière (Dordogne)." In *Les Industries à pointes foliacées du Paléolithique supérieur européen*, edited by J. Kozlowski, 293–320. ERAUL, No. 42, Liège.

Geribas, Núria, Marina Mosquera, and Josep Maria Vergès. 2010. "What Novice Knappers Have to Learn to Become Expert Stone Toolmakes." *Journal of Archaeological Science* 37: 2857–2870.

Gurtov, Alia N., and Metin I. Eren. 2014. "Lower Paleolithic Bipolar Reduction and Hominin Selection of Quartz at Olduvai Gorge, Tanzania: What's the Connection?" *Quaternary International* 322–323: 285–291.

Gurtov, Alia N., Briggs Buchanan, and Metin I. Eren. 2015. "'Dissecting' Basalt and Quartzite Bipolar Flake Shape: A Morphometric Comparison of Experimental Replications from Olduvai Gorge, Tanzania." *Lithic Technology* 40: 332–341.

Hay, Conran A. 1978. "Comment on A History of Flint-Knapping Experimentation, 1838–1976." *Current Anthropology* 19: 360–361.

Hayashi, Kensaku. 1968. "The Fukui Microblade Technology and Its Relationships in Northeast Asia and North America." *Arctic Anthropology* 5: 128–190.

Hiscock, Peter, and Chris Clarkson. 2005. "Experimental Evaluation of Kuhn's Geometric Index of Reduction and the Flat-Flake Problem." *Journal of Archaeological Science* 32: 1015–1022.

Hunzicker, David A. 2008. "Folsom Projectile Technology: An Experiment in Design, Effectiveness and Efficiency." *Plains Anthropologist* 53: 291–311.

Horowitz, Rachel A., and Grant S. McCall. 2013. "Evaluating Indices of Curation for Archaic North American Bifacial Projectile Points." *Journal of Field Archaeology* 38: 347–361.

Huckell, Bruce B. 1982. "The Denver Elephant Project: A Report on Experimentation with Thrusting Spears." *Plains Anthropologist* 27: 217–224.

Iovita, Radu, Holger Schönekeß, Sabine Gaudzinski-Windheuser, and Frank Jäger. 2014. "Projectile Impact Fractures and Launching Mechanisms: Results of Controlled Ballistic Experiment using Replica Levallois Points." *Journal of Archaeological Science* 48: 73–83.

Jennings, Thomas A. 2011. "Experimental Production of Bending and Radial Flake Fractures and Implications for Lithic Technologies." *Journal of Archaeological Science* 38: 3644–3651.

Jennings, Thomas A., Charlotte D. Pevny, and William A. Dickens. 2010. "A Biface and Blade Core Efficiency Experiment: Implications for Early Paleoindian Technological Organization." *Journal of Archaeological Science* 37: 2155–2164.

Johnson, L. L. 1978. "A History of Flint-Knapping Experimentation, 1838–1976." *Current Anthropology* 19: 337–372.

de Juana, S., A.B. Galán, and M. Domínguez-Rodrigo. 2010. "Taphonomic Identification of Cut Marks Made with Lithic Handaxes: An Experimental Study." *Journal of Archaeological Science* 37: 1841–1850.

Katz, Paul R. 1978. "Comment on A History of Flint-Knapping Experimentation, 1838–1976." *Current Anthropology* 19: 362.

Kelly, Robert L. 1994, "Some Thoughts on Future Directions in the Study of Stone Tool Organization." In *The Organization of North American Prehistoric Chipped Stone Tool Technology*, edited by Phillip Carr, 132–136. Ann Arbor: International Monographs in Prehistory 7.

Kelterborn, P. 1999. "Analysen und Experimente zu Herstellung und Gebrauch von Horgener Pfeilspitzen." Jahrbuch der Schweizerischen Gesellschaft für Ur- und Frühgeschichte 83: 37–64.

Kelterborn, Peter. 2000. "Analysen und Experimente zu Herstellung und Gebrauch von Horgener Pfeilspitzen." *Jahrbuch der Schweizerischen Gesellschaft für Ur-und Frühgeschichte* 83: 37–64.

Key, Alastair J.M. 2013. "Applied Force as a Determining Factor in Lithic Use-Wear Accrual: An Experimental Investigation of Its Validity as a Method with Which to Infer Hominin Upper Limb Biomechanics." *Lithic Technology* 38: 32–45.

Key, Alastair J.M., and Christopher J. Dunmore. 2015. "The Evolution of the Hominin Thumb and the Influence Exerted by the Non-Dominant Hand During Stone Tool Production." *Journal of Human Evolution* 78: 60–69.

Key, Alastair J.M., and Stephen J. Lycett. 2011. "Technology Based Evolution? A Biometric Test of the Effects of Handsize versus Tool form on Efficiency in an Experimental Cutting Task." *Journal of Archaeological Science* 38: 1663–1670.

——. 2014. "Are Bigger Flakes Always Better? An Experimental Assessment of Flake Size Variation on Cutting Efficiency and Loading." *Journal of Archaeological Science* 41: 140–146.

——. 2015. "Edge Angle as a Variably Influential Factor in Flake Cutting Efficiency: An Experimental Investigation of its Relationship with Tool Size and Loading." *Archaeometry* 57: 911–927.

Knudson, Ruthann. 1978. "Comment on A History of Flint-Knapping Experimentation, 1838–1976." *Current Anthropology* 19: 362–363.

Kuhn, Steven L. 1990. "A Geometric Index of Reduction for Unifacial Stone Tools." *Journal of Archaeological Science* 17: 583–593.

Laughlin, John P., and Robert L. Kelly. 2010. "Experimental Analysis of the Practical Limits of Lithic Refitting." *Journal of Archaeological Science* 37: 427–433.

Lemorini, Cristina, Thomas W. Plummer, David R. Braun, Alyssa N. Crittenden, Peter W. Ditchfield, Laura C. Bishop, Fritz Hertel, *et al.* 2014. "Old Stones' Song: Use-Wear Experiments and Analysis of the Oldowan Quartz and Quartzite Assemblage from Kanjera South (Kenya)." *Journal of Human Evolution* 72: 10–25.

Lerner, Harry. 2013. "Experiments and Their Application to Lithic Archaeology: An Experimental Essay." In *Human Expeditions: Inspired by Bruce Trigger*, edited by Stephen Chrisomalis, and Andre Costopoulos, 73–89. Toronto: University of Toronto Press.

Lerner, Harry, Xiangdong Du, Andre Costopoulos, and Martin Ostoja-Starzewski. 2007. "Lithic Raw Material Physical Properties and Use-Wear Accrual." *Journal of Archaeological Science* 34: 711–722.

Letourneux, Claire, and Jean-Marc Pétillon. 2008. "Hunting Lesions Caused by Osseous Projectile Points: Experimental Results and Archaeological Implications." *Journal of Archaeological Science* 35: 2849–2862.

Lett, James. 1997. *Science, Reason, and Anthropology: The Principles of Rational Inquiry.* New York: Rowman & Littlefield Publishers, Inc.

Li, Feng. 2015. "An Experimental Study of Bipolar Reduction at Zhoukoudian Locality 1, North China." *Quaternary International* 400: 23–29.

Lipo, Carl P., Robert C. Dunnell, Michael J. O'Brien, Veronica Harper, and John Dudgeon. 2012. "Beveled Projectile Points and Ballistics Technology." *American Antiquity* 77: 774–788.

Lombard, Marlize, Isabelle Parsons, and M. M. van der Ryst. 2004. "Middle Stone Age Lithic Point Experimentation for Macro-Fracture and Residue Analyses: The Process and Preliminary Results with Reference to Sibudu Cave Points." *South African Journal of Science* 100: 159–166.

Lombard, Marlize, and Justin Pargeter. 2008. "Hunting with Howiesons Poort Segments: Pilot Experimental Study and the Functional Interpretation of Archaeological Tools." *Journal of Archaeological Science* 35: 2523–2531.

Lycett, Stephen J., and Parth Chauhan. 2010. "Analytical Approaches to Palaeolithic Technologies: An Introduction." In *New Perspectives on Old Stones*, edited by Stephen J. Lycett, and Parth Chauhan, 1–22. New York: Springer.

Lycett, Stephen J., and Metin I. Eren. 2013a. "Levallois Lessons: The Challenge of Integrating Mathematical Models, Quantitative Experiments and the Archaeological Record." *World Archaeology* 45: 519–538.

——. 2013b. "Levallois Economics: An Examination of 'Waste' Production in Experimentally Produced Levallois Reduction Sequences." *Journal of Archaeological Science* 40: 2384–2392.

Mackay, Alex. 2008. "A Method for Estimating Edge Length from Flake Dimensions: Use and Implications for Technological Change in the Southern African MSA." *Journal of Archaeological Science* 35: 614–622.

MacDonald, Danielle A. 2014. "The Application of Focus Variation Microscopy for Lithic Use-Wear Quantification." *Journal of Archaeological Science* 48: 26–33.

Machin, A.J., R.T. Hosfield, and S.J. Mithen. 2007. "Why are Some Handaxes Symmetrical? Testing the Influence of Handaxe Morphology on Butchery Effectiveness." *Journal of Archaeological Science* 34: 883–893.

Magnani, Matthew, Zeljko Rezek, Sam C. Lin, Annie Chan, and Harold L. Dibble. 2014. "Flake Variation in Relation to the Application of Force." *Journal of Archaeological Science* 46: 37–49.

Magne, Martin P.R. 1985. *Lithics and Livelihood: Stone Tool Technologies of Central and Southern Interior B.C.* Ottawa: Archaeology Survey of Canada, Mercury Series No. 133.

Mahaney, Robert A. 2014. "Exploring the Complexity and Structure of Acheulean Stoneknapping in Relation to Natural Language." *Paleoanthropology* 2014: 586–606.

Malik, S.C. 1978. "Comment on A History of Flint-Knapping Experimentation, 1838–1976." *Current Anthropology* 19: 364.

Marsh, Erik J., and Jeffrey R. Ferguson. 2010. "Introduction." In *Designing Experimental Research in Archaeology: Examining Technology Through Production and Use*, edited by Jeffrey R. Ferguson, 1–13. Boulder: University Press of Colorado.

Marwick, Ben. 2008. "What Attributes Are Important for the Measurement of Assemblage Reduction Intensity? Results from an Experimental Stone Artifact Assemblage with Relevance to the Hoabinhian of Mainland Southeast Asia." *Journal of Archaeological Science* 35: 1189–1200.

McCall, Grant S., and Brent Pelton. 2010. "The Iceman Knappeth: Learning About Soft Stones from an Experiment Knapping Ice." In *Pushing the Envelope: Experimental Directions in the Archaeology of Stone Tools*, edited by Grant McCall, 103–110. New York: Nova Science Publishers, Inc.

McPherron, Shannon P., David R. Braun, Tamara Dogandžića, Will Archer, Dawit Desta, and Sam C. Lin. 2014. "An Experimental Assessment of the Influences on Edge Damage to Lithic Artifacts: A Consideration of Edge

Angle, Substrate Grain Size, Raw Material Properties, and Exposed Face." *Journal of Archaeological Science* 49: 70–82.

Mercieca, Alizon, and Peter Hiscock. 2008. "Experimental Insights into Alternative Strategies of Lithic Heat Treatment." *Journal of Archaeological Science* 35: 2634–2639.

Mesoudi, Alex. 2011. *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences*. Chicago: University of Chicago Press.

Miller, G. Logan. 2015. "Ritual Economy and Craft Production in Small-Scale Societies: Evidence from the Microwear Analysis of Hopewell Bladelets." *Journal of Anthropological Archaeology* 39: 124–138.

Morales, Juan I., Carlos Lorenzo, and Josep M. Vergès. 2015. "Measuring Retouch Intensity in Lithic Tools: A New Proposal Using 3D Scan Data." *Journal of Archaeological Method and Theory* 22: 543–558.

Morgan, Brooke, Metin I. Eren, Nada Khreisheh, Genevieve Hill, and Bruce Bradley. 2015a. "Clovis Bipolar Lithic Reduction at Paleo Crossing, Ohio: A Reinterpretation Based on the Examination of Experimental Replications." In *Clovis: On the Edge of a New Understanding*, edited by Thomas Jennings, and Ashley Smallwood, 121–144. College Station: Texas A&M Press.

Morgan, T. J. H., N. T. Uomini, L. E. Rendell, L. Chouinard-Thuly, S. E. Street, H. M. Lewis, C. P. Cross, *et al.* 2015b. "Experimental Evidence for the Co-Evolution of Hominin Tool-Making Teaching and Language." *Nature Communications* 6: 1–8.

Morrow, Julie E. 1997. "End Scraper Morphology and Use-Life: An Approach for Studying Paleoindian Lithic Technology and Mobility." *Lithic Technology* 22: 70–85.

Moss, E.H., Newcomer, M.H., 1982. "Reconstruction of tool use at Pincevent: Microwear and experiments." In Tailler! pour quoi faire: Prehistoire et technologie lithique II - Recent Progression in Microwear studies, edited by Daniel Cahen, 289–346. Koninklijk Museum voor Midden-Afrika, Tervuren.

Muller-Beck, Hansjürgen. 1978. "Comment on A History of Flint-Knapping Experimentation, 1838–1976." *Current Anthropology* 19: 364.

Nami, Hugo G. 2010. "Theoretical Reflections on Experimental Archaeology and Lithic Technology: Issues on Actualistic Stone Tools Analysis and Interpretation." In *Experiments and Interpretation of Traditional Technologies: Essays in Honor of Errett Callahan*, edited by Hugo Nami, 91–168. Buenos Aires: Ediciones de Arqueología Contemporánea.

Nigra, Benjamin T., and Jeanne E. Arnold. 2013. "Explaining the Monopoly in Shell-Bead Production on the Channel Islands: Drilling Experiments with Four Lithic Raw Materials." *Journal of Archaeological Science* 40: 3647–3659.

Nonaka, Tetsushi, Blandine Bril, and Robert Rein. 2010. "How do Stone Knappers Predict and Control the Outcome of Flaking? Implications for Understanding Early Stone Tool Technology." *Journal of Human Evolution* 59: 155–167.

O'Brien, Michael J. 2010. "The Future of Paleolithic Studies: A View from the New World." In *New Perspectives on Old Stones*, edited by Stephen J. Lycett, and Parth Chauhan, 311–334. New York: Springer.

Odell, George H., and Frank Cowan. 1986. "Experiments with Spears and Arrows on Animal Targets." *Journal of Field Archaeology* 13: 195–212.

O'Farrell, M., 2004. "Les pointes de La Gravette de Corbiac (Dordogne) et considérations sur la chasse au Paléolithique supérieur ancien." In Approches Fonctionnelles en Préhistoire edited by Pierre Bodu and Claude Constantin 121–138. Société Préhistorique Française, Paris.

Olausson, Deborah. 2010. "Experimental Flintknapping Replication: A Valuable Method of Archaeological Analysis." In *Experiments and Interpretation of Traditional Technologies: Essays in Honor of Errett Callahan*, edited by Hugo Nami, 37–55. Buenos Aires: Ediciones de Arqueología Contemporánea.

Outram, Alan K. 2008. "Introduction to Experimental Archaeology." *World Archaeology* 40: 1–6.

Owen, Richard. 1835. "XXIII. On the Anatomy of the *Calyptræidæ*." *The Transactions of the Zoological Society of London* 1: 207–212.

Pargeter, Justin. 2011. "Assessing the Macrofracture Method for Identifying Stone Age Hunting Weaponry." *Journal of Archaeological Science* 38: 2882–2888.

——. 2013. "Rock Type Variability and Impact Fracture Formation: Working Towards a more Robust Macrofracture Method." *Journal of Archaeological Science* 40: 4056–4065.

Pargeter, Justin, and Justin Bradfield. 2012. "The Effects of Class I and II Sized Bovids on Macrofracture Formation and Tool Displacement: Results of a Trampling Experiment in a Southern African Stone Age context." *Journal of Field Archaeology* 37: 238–251.

Patten, Robert J. 2002. "Solving the Folsom Fluting Problem." In *Folsom Technology and Lifeways*, edited by John Clark, and Michael Collins, 299–308. Tulsa: Special Publication No. 4, Lithic Technology.

———. 2005. *Peoples of the Flute: A Study in Anthropolithic Forensics*. Lakewood: Stone Dagger Publications.

———. 2009. *Old Tools — New Eyes: A Primal Primer of Flintknapping*. Lakewood: Stone Dagger Publications.

Pelcin, Andrew. 1997a. "The Effect of Indentor Type on Flake Attributes: Evidence from a Controlled Experiment." *Journal of Archaeological Science* 24: 613–621.

———. 1997b. "The Effect of Core Surface Morphology on Flake Attributes: Evidence from a Controlled Experiment." *Journal of Archaeological Science* 24: 749–756.

———. 1997c. "The Formation of Flakes: The Role of Platform Thickness and Exterior Platform Angle in the Production of Flake Initiations and Terminations." *Journal of Archaeological Science* 24: 1107–1113.

———. 1998. "The Threshold Effect of Platform Width: A Reply to Davis and Shea." *Journal of Archaeological Science* 25: 615–620.

Pétillon, Jean-Marc, Olivier Bignon, Pierre Bodu, Pierre Cattelain, Grégory Debout, Mathieu Langlais, Véronique Laroulandie, Hugues Plisson, and Boris Valentin. 2011. "Hard Core and Cutting Edge: Experimental Manufacture and use of Magdalenian Composite Projectile Tips." *Journal of Archaeological Science* 38: 1266–1283.

Pettigrew, Devin B., John C. Whittaker, Justin Garnett, and Patrick Hashman. 2015. "How Atlatl Darts Behave: Beveled Points and the Relevance of Controlled Experiments." *American Antiquity* 80: 590–601.

Pevny, Charlotte D. 2012. "Distinguishing Taphonomic Processes from Stone Tool Use at the Gault Site, Texas." In *Contemporary Lithic Analysis in the Southeast: Problems, Solutions, and Interpretations*, edited by Philip Carr, Andrew Bradbury, and Sarah Price, 55–78. Tuscaloosa: The University of Alabama Press.

Plisson, Hughes, and Sylvia Beyries. 1998. "Pointes ou outils triangulaires? Données fonctionnelles dans le Moustérien Levantin. Paléorient." *Interdisciplinary Review of Prehistory and Protohistory of Southwestern Asia* 24: 5–24.

Prasciunas, Mary. 2007. "Bifacial Cores and Flake Production Efficiency: An Experimental Test of Technological Assumptions." *American Antiquity* 72: 334–348.

Presnyakova, Darya, Will Archer, David R. Braun, and Wesley Flear. 2015. "Documenting Differences Between Early Stone Age Flake Production Systems: An Experimental Model and Archaeological Verification." *PLoS ONE* 10: e0130732.

Price, Sarah E. 2012. "Omnipresent? We don't Recover the Half of it!." In *Contemporary Lithic Analysis in the Southeast: Problems, Solutions, and Interpretations*, edited by Philip Carr, Andrew Bradbury, and Sarah Price, 13–27. Tuscaloosa: The University of Alabama Press.

Putt, Shelby S., Alexander D. Woods, and Robert G. Franciscus. 2014. "The Role of Verbal Interaction During Experimental Bifacial Stone Tool Manufacture." *Lithic Technology* 39: 96–112.

Putt, Shelby S. 2015. "The Origins of Stone Tool Reduction and the Transition to Knapping: An Experimental Approach." *Journal of Archaeological Science: Reports* 2: 51–60.

Quinn, Colin P., William Andrefsky, Jr., Ian Kuijt, and Bill Finlayson. 2009. "Perforation with Stone Tools and Retouch Intensity: A Neolithic Case Study." In *Lithic Technology: Measures of Production, Use, and Curation*, edited by William Andrefsky, Jr., 150–174. Cambridge: University of Cambridge Press.

Ranere, Anthony J. 1978. "Comment on A History of Flint-Knapping Experimentation, 1838–1976." *Current Anthropology* 19: 365.

Rasic, Jeffrey, and William Andrefsky, Jr. 2001. "Alaskan Blade Cores as Specialized Components of Mobile Toolkits: Assessing Design Parameters and Toolkit Organization Through Debitage Analysis." In *Lithic Debitage: Context, Form, Meaning*, edited by William Andrefsky, Jr., 61–79. Cambridge: University of Cambridge Press.

Rezek, Zeljko, Sam Lin, Radu Iovita, and Harold L. Dibble. 2011. "The Relative Effects of Core Surface Morphology on Flake Shape and other Attributes." *Journal of Archaeological Science* 38: 1346–1359.

Reti, Jay S. 2014. Methods for Determining Differential Behaviors in Stone Tool Production and Application to the Oldowan of Olduvai Gorge, Tanzania and Koobi Fora, Kenya. Rutgers University, New Brunswick, NJ. Ph.D. Dissertation.

Rodríguez-Rellán, Carlos, Ramón Fábregas Valcarce, and Elías Berriochoa Esnaola. 2013. "Shooting out the Slate: Working with Flaked Arrowheads Made on Thin-Layered Rocks." *Journal of Archaeological Science* 38: 1939–1948.

Roe, Brian E., and David R. Just. 2009. "Internal and External Validity in Economics Research: Tradeoffs Between Experiments, Field Experiments, Natural Experiments, and Field Data." *American Journal of Agricultural Economics* 91: 1266–1271.

Rolian, Campbell, Daniel E. Lieberman, and John P. Zermeno. 2011. "Hand Biomechanics During Simulated Stone Tool Use." *Journal of Human Evolution* 61: 26–41.

Rots, Veerle, and Hughes Plisson. 2014. "Projectiles and the Abuse of the Use-Wear Method in a Search for Impact." *Journal of Archaeological Science* 48: 154–165.

Roux, V., B. Bril, and G. Dietrich. 1995. "Skills and Learning Difficulties Involved in Stone Knapping: The Case of Stone-Bead Knapping in Khambhat, India." *World Archaeology* 27: 63–87.

Sano, Katsuhiro. 2009. "Hunting Evidence from Stone Artefacts from the Magdalenian Cave Site Bois Laiterie, Belgium: A Fracture Analysis." *Quartar* 56: 67–86.

Sano, Katsuhiro, and Masayoshi Oba. 2014. "Projectile Experimentation for Identifying Hunting Methods with replicas of Upper Palaeolithic Weaponry from Japan." In *International Conference on Use-Wear Analysis: Use-Wear 2012*, edited by N.B. João Marreiros, and J.G. Bao, 466–478. Newcastle upon Tyne: Cambridge Scholars Publishing.

Schillinger, Kerstin, Alex Mesoudi, and Stephen J. Lycett. 2014a. "Copying error and the Cultural Evolution of "additive" vs. "reductive" Material Traditions: An Experimental Assessment." *American Antiquity* 79: 128–143.

——. 2014b. "Considering the Role of Time Budgets on Copy-Error Rates in Material Culture Traditions: An Experimental Assessment." *PLoS ONE* 9: e97157.

——. 2015. "The Impact of Imitative versus Emulative Learning Mechanisms on Artifactual Variation: Implications for the Evolution of Material Culture." *Evolution and Human Behavior* 36: 446–455.

Schindler, Bill, and Jeremy Koch. 2012. "Flakes Giving You Lip? Let Them Speak: An Examination of the Relationship Between Percussor Type and Lipped Platforms." *Archaeology of Eastern North America* 40: 99–106.

Schmidt, Patrick, Sylvie Masse, Guillaume Laurent, Aneta Slodczyk, Eric Le Bourhis, Christian Perrenoud, Jacques Livage, and François Fröhlich. 2012. "Crystallographic and Structural Transformations of Sedimentary Chalcedony in Flint upon Heat Treatment." *Journal of Archaeological Science* 39: 135–144.

Schmidt, Patrick, V. Lea, Ph. Sciau, and F. Fröhlich. 2013. "Detecting and Quantifying Heat Treatment of Flint and other Silica Rocks: A New Non-Destructive Method Applied to Heat-Treated Flint from the Neolithic Chassey Culture, Southern France." *Archaeometry* 55: 794–805.

Shea, John, Zachary Davis, and Kyle Brown. 2001. "Experimental tests of Middle Palaeolithic Spear Points using a Calibrated Crossbow." *Journal of Archaeological Science* 28: 807–816.

Shea, J.J., Brown, K., Davis, Z., 2002. "Controlled experiments with Middle Paleolithic spear points: Levallois points." In *Experimental Archaeology: Replicating Past Objects, Behaviors, and Processes*. Edited by James R. Mathieu, 55–72. British Archaeological Reports: Oxford.

Shea, John J. 2015. "Making and using stone tools: advice for learners and teachers and insights for archaeologists." *Lithic Technology* 40: 231–248.

Shelley, Phillip H. 1990. "Variation in Lithic Assemblages: An Experiment." *Journal of Field Archaeology* 17: 187–193.

Shipton, Ceri, Michael D. Petraglia, and K. Paddayya. 2009. "Stone Tool Experiments and Reduction Methods at the Acheulean Site of Isampur Quarry, India." *Antiquity* 83: 769–785.

Short, Michael J., Hunzicker, D. A., Patten, Bob. 2007. "Pattern and allometric measurement of reduction in experimental Folsom bifaces." *Lithic Technology* 32: 203–217.

Shott, Michael J. 2002. "Weibull Estimation on use Life Distribution in Experimental Spear-Point Data." *Lithic Technology* 27: 93–109.

Shott, Michael J., Andrew P. Bradbury, Philip J. Carr, and George H. Odell. 2000. “Flake Size from Platform Attributes: Predictive and Empirical Approaches.” *Journal of Archaeological Science* 27: 877–894.

Shott, Michael J., and Kathryn J. Weedman. 2007. “Measuring Reduction in Stone Tools: An Ethnoarchaeological Study of Gamo Hidescrapers from Ethiopia.” *Journal of Archaeological Science* 34: 1016–1035.

Sisk, Matthew L., and John Shea. 2009. “Experimental Use and Quantitative Performance Analysis of Triangular Flakes (Levallois points) Used as Arrowheads.” *Journal of Archaeological Science* 36: 2039–2047.

Smallwood, Ashley M. 2013. “Building Experimental Use-Wear Analogues for Clovis Biface Functions.” *Archaeological and Anthropological Sciences* 7: 13–26.

Sollberger, J. B., and L. W. Patterson. 1976. “Prismatic Blade Replication.” *American Antiquity* 41: 517–531.

Soriano, S., 1998. “La production de lamelles torses dans le niveaux du Paléolithique supérieur ancien d'Umm El Tlel (Syrie): exploration théoretique et experimentale de ses modalités.” In Préhistoire d'Anatolie: Genèse de deux mondes/Anatolian Prehistory: At the Crossroad of Two Worlds, Edited by Marcel Otte, 731–748. Etudes et Recherches Archéologiques de L'Université de Liège, Liège.

Stafford, Michael. 2003. “The Parallel-Flaked Flint Daggers of Late Neolithic Denmark: An Experimental Perspective.” *Journal of Archaeological Science* 30: 1537–1550.

Stout, Dietrich, Nick Toth, Kathy Shick, J. Stout, and G. Hutchins. 2000. “Stone Tool-Making and Brain Activation: Position Emission Tomography (PET) Studies.” *Journal of Archaeological Science* 27: 1215–1223.

Stout, D., Semaw, S. (2006). “Knapping skill of the earliest stone toolmakers: insights from the study of modern human novices. In The Oldowan: Case Stu207dies into the Earliest Stone Age, edited by Nicholas Toth and Kathy Schick, pp. 307–320.” Stone Age Institute Publication Series, Bloomington.

Stout, D., Apel, J., Commander, J., Roberts, M. 2014. “Late Acheulean technology and cognition at Boxgrove, UK.” *Journal of Archaeological Science* 41: 576–590.

Surovell, Todd A. 2009. *Toward a Behavioral Ecology of Lithic Technology: Cases from Paleoindian Archaeology.* Tucson: University of Arizona Press.

Tactikos, Joanne C. 2003. “A Re-Evaluation of Palaeolithic Stone Tool Cutting Edge Production Rates and their Implications.” In *Lithic Analysis at the Millennium*, edited by Nora Moloney, and Michael Shott, 151–162. Walnut Creek, CA: Left Coast Press.

Tallavaara, Miikka, Mikael A. Manninen, Esa Hertell, and Tuija Rankama. 2010. “How Flakes Shatter: A Critical Evaluation of Quartz Fracture Analysis.” *Journal of Archaeological Science* 37: 2442–2448.

Temple, Carolyn R., and Robert Lee Sappington. 2013. “An Experimental Archaeological Study of the Effects of Off-Road Vehicles on Lithic Scatters.” *Journal of Northwest Anthropology* 47: 167–187.

Thomas, David Hurst. 1986. “Points on Points: A Reply to Flenniken and Raymond.” *American Antiquity* 51: 619–627.

Titmus, Gene L., and James Woods. 1986. “An Experimental Study of Projectile Point Fracture Patterns.” *Journal of California and Great Basin Anthropology* 8: 37–39.

Tryon, Christian A., Sally McBrearty, and Pierre-Jean Texier. 2005. “Levallois Lithic Technology from the Kapthurin Formation, Kenya: Acheulian Origin and Middle Stone Age Diversity.” *African Archaeological Review* 22: 199–229.

Uomini, Natalie T., and Georg F. Meyer. 2013. “Shared Brain Lateralization Patterns in Language and Acheulean Stone Tool Production: A Functional Transcranial Doppler Ultrasound Study.” *PLoS ONE* 8: e72693.

Wadley Lyn, Prinsloo Linda C. 2014. “Experimental Heat Treatment of Silcrete Implies Analogical Reasoning in the Middle Stone Age.” *Journal of Human Evolution* 70: 49–60.

Waguespack, Nicole M., Todd A. Surovell, Allen Denoyer, Alice Dallow, Adam Savage, Jamie Hyneman, and Dan Tapster. 2009. “Making a Point: Wood-Versus Stone-Tipped Projectiles.” *Antiquity* 83: 786–800.

Weitzel, Celeste. 2010. El estudio de los artefactos formatizados fracturados. Contribución a la comprensión del registro arqueológico y las actividades humanas, Unpublished Ph.D. dissertation. Universidad de Buenos Aires, Argentina.

Weitzel, Celeste, Karen Borrazzo, Antonio Ceraso, and Catalina Balirán. 2014a. "Trampling Fragmentation Potential of Lithic Artifacts: An Experimental Approach." *Intersecciones en antropología Taphonomic Approaches to the Archaeological Record* 15:97–110.

Weitzel, Celeste, Nora Flegenheimer, Mariano Colombo, and Jorge Martinez. 2014b. "Breakage Patterns on Fishtail Projectile Points: Experimental and Archaeological Cases." *Ethnoarchaeology* 6: 81–102.

Wenban-Smith, F. F. 1989. "The use of Canonical Variates for Determination of Biface Manufacturing Technology at Boxgrove Lower Palaeolithic Site and the Behavioural Implications of This Technology." *Journal of Archaeological Science* 16: 17–26.

West, J. A., Louys, J. 2007. "Differentiating bamboo from stone tool cut marks in the zooarchaeological record, with a discussion on the use of bamboo knives." *Journal of Archaeological Science* 34: 512–518.

Whittaker, John C., and Grant McCall. 2001. "Handaxe-hurling Hominids: An Unlikely Story." *Current Anthropology* 42: 566–572.

Whittaker, John C. 1987. "Individual Variation as an Approach to Economic Organization: Projectile Points at Grasshopper Pueblo, Arizona." *Journal of Field Archaeology* 14: 465–479.

——. 1994. *Flintknapping:Making and Understanding Stone Tools*. Austin: University of Texas Press.

——. 2004. *American Flintknappers: Stone Age Art in the Age of Computers*. Austin: University of Texas Press.

Wilkins, Jayne, Benjamin J. Schoville, and Kyle S. Brown. 2014. "An Experimental Investigation of the Functional Hypothesis and Evolution Advantage of Stone-Tipped Spears." *PLoS ONE* 9: e104514.

Wilkins, Jayne, Benjamin J. Schoville, Kyle S. Brown, and Michael Chazan. 2012. "Evidence for Early Hafted Hunting Technology." *Science* 338: 942–946.

——. 2015. "Kathu Pan 1 points and the Assemblage-Scale, Probabilistic Approach: A Response to Rots and Plisson, "Projectiles and the Abuse of the Use-Wear Method in a Search for Impact"." *Journal of Archaeological Science* 54: 294–299.

Williams, Erin Marie, Adam D. Gordon, and Brian G. Richmond. 2012. "Hand Pressure Distribution During Oldowan Stone Tool Production." *Journal of Human Evolution* 62: 520–532.

Williams, Erin M.; Gordon, A. D.; Richmond, Brian G. 2014. "Biomechanical strategies for accuracy and force generation during stone tool production." Journal of Human Evolution 72: 52–63.

Williams, Justin P., and William Andrefsky, Jr. 2011. "Debitage Variability Among Multiple Flint Knappers." *Journal of Archaeological Science* 38:865–872.

Wilmsen, Edwin N. 1970. Lithic Analysis and Cultural Inference: a Paleoindian Case. University of Arizona Press.

Wilson, J., Andrefsky Jr, W. 2008. "In Lithic Technology." In *Exploring Retouch on Bifaces: Unpacking Production, Resharpening, and Hammer Type*, edited by W. Andrefsky, Jr., pp. 86–105. Cambridge University Press.

Winton, Vicky. 2005. "An Investigation of Knapping-Skill Development in the Manufacture of Palaeolithic Handaxes." In *Stone Knapping: The Necessary Conditions for a Uniquely Hominin Behavior*, edited by Valentine Roux, and Blandine Bril, 109–116. Oxford: McDonald Institute for Archaeological Research.

Yaroshevich, A., Kaufman, D., Nuzhnyy, D., Bar-Yosef, O., Weinstein-Evron, M., 2010. "Design and performance of microlith implemented projectiles during the Middle and the Late Epipaleolithic of the Levant: experimental and archaeological evidence." *Journal of Archaeological Science* 37, 368–388.

Notes on contributors

Metin I. Eren is Assistant Professor of Anthropology at Kent State University, Kent, Ohio, U.S.A. He is a corresponding author for this paper.

Correspondence to: Metin I. Eren. Email: meren@kent.edu

Stephen J. Lycett is Associate Professor of Anthropology at the University at Buffalo, SUNY. He is the co-corresponding author for this paper. Stephen J. Lycett. Email: sjlycett@buffalo.edu

Robert J. Patten is a B.S. Graduate of Civil Engineering at Colorado State University. Recipient of the Crabtree award from the Society of American Archaeology in 2004, his research interests focus on process controls used in lithic technology.

Briggs Buchanan is an Assistant Professor of Anthropology at the University of Tulsa, Tulsa, Oklahoma, U.S.A.

Justin Pargeter is a PhD candidate at Stony Brook University, New York, and a Senior Research Fellow at the University of Johannesburg, South Africa.

Michael J. O'Brien is a Professor of Anthropology, University of Missouri, Columbia, Missouri, U.S.A.