Distribution Agreement

Jana Muschinski April 15, 2015

Reduction in the Acheulean:

Identification of a relative reduction equation and application to the Boxgrove assemblage

by

Jana Muschinski

Dietrich Stout, Ph.D.
Adviser

Department of Anthropology

Dietrich Stout, Ph.D.

Adviser

Liv Nilsson Stutz, Ph.D.

Committee Member

Hilary Gopnik, Ph.D.

Committee Member

2015

Reduction in the Acheulean:

Identification of a relative reduction equation and application to the Boxgrove assemblage

By

Jana Muschinski

Dietrich Stout, Ph.D.

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Anthropology

2015

Abstract

Reduction in the Acheulean:

Identification of a relative reduction equation and application to the Boxgrove assemblage
By Jana Muschinski

One of the most common sources of information regarding early hominid behavior is stone tool debitage. Flakes are made during the creation of a stone tool, and if these flakes are viewed chronologically, they can be represented as a reduction sequence. Being able to identify what portion of the reduction sequence is present at an archaeological site would provide valuable information about resource transportation and management. Several studies have attempted to create a percent completion model for bifacial lithic tool creation based on flake attributes (Bradbury & Carr, 1999; Ingbar et al., 1989; Shott, 1996). This study creates a completion proportion equation for refined Acheulean handaxe debitage assemblages of flint using multiple linear regression. The identified model has an adjusted R-squared value of 0.434 and achieves a slope of 1 when the predicted values are plotted against the actual reduction proportion values for each flake.

This model was applied to data from the middle Pleistocene Acheulean site of Boxgrove. According to the model, the assemblage consists of primarily early reduction flakes, contrary to past interpretations (Roberts & Parfitt, 1999). However, this application is problematic because the created model does not fully control for size, which differs between assemblages. Future research on both the model and its application to Boxgrove and other sites will be crucial. Ideally, the sample size used to create the models could be increased through further experimental replications and the model could be tested upon a further experimental assemblage. The overall effects of using a size dependent variable within the reduction equation also need to be investigated further. On the broadest scale, such reduction models could ideally be created for a variety of materials and technologies, enabling application to a wide range of archaeological sites.

Reduction in the Acheulean:

Identification of a relative reduction equation and application to the Boxgrove assemblage

By

Jana Muschinski

Dietrich Stout, Ph.D.

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Anthropology

2015

Acknowledgements

I greatly appreciate the support and guidance of Dr. Stout, and the encouragement and feedback of my committee members Dr. Stutz and Dr. Gopnik. Completing this thesis has helped me develop skills that will be very useful in graduate school and throughout the remainder of my life.

I would additionally like to thank my friends and family members who have read drafts and helped me in the editing process.

Table of Contents

Table of Figures and Tables

Introduction

A large part of what we know today about the past, especially the distant past, relies on an understanding of ancient technologies. When it comes to our earliest tool using ancestors, well preserved evidence becomes very scarce. Unlike organic materials, stone preserves well and relatively easily. For this reason, debitage assemblages, the materials created during the flaking of stone, are often among the Paleolithic finds richest in information.

The range of behavior that lithics can provide information about is simultaneously expansive and limited. The creation and use of stone tools involves resource collection, management, distribution, and transportation, decision making, forethought, and an understanding of cause and effect. Stone tool wear and tear can provide clues about hunting, scavenging, butchering, or gathering methods, thereby providing information about daily activities (Stemp et al., 2009). The study of stone tools may even provide clues about the history of language (Stout & Chaminade, 2012). By learning more about Paleolithic hominins, we learn more about the history of humanity and what it means to be human. While the potential is great, in practice studies are limited by a number of factors.

In the study of resource transport, one could ideally trace material across a landscape. By tracing the movement of stone through a region, starting at the raw-material source and moving through all following modifications and uses of the material, it would become possible to build a picture of the processes and decisions that went into tool creation and use (Braun et al., 2008). By having a more detailed understanding of the distribution of materials, one could make conclusions about raw material economy, resource transportation, resource management, and overall decision making skills (Braun et al., 2008). Tool-related behavior can then act as a window into earlier hominin behavior and cognition. However, for this to work, we need to find a way to identify what part of the reduction sequence is represented within an assemblage at a site. Ideally, it would be possible to develop a percent completion equation using several measurable flake attributes that, when applied, would give you a prediction of where a flake falls within the overall tool production process. You could thereby analyze all flakes in a site assemblage, establishing what portion of tool production occurred there. Braun et al. used calculations of reduction intensity to successfully investigate tool production and transport behaviors in the Developed Oldowan of Koobi Fora, relating reduction intensity to distance from raw-material sources (2008). While several studies have identified potentially useful models for Acheulean technology, a reliably applicable and well-controlled model does not yet exist.

A General Background

A Brief History of the Study of Lithics

Until the late 1700s, it was considered implausible for early human ancestors to have been creating and using stone tools (Kooyman, 2000). As more archaeological sites were excavated and archaeological theory and methods changed, interest in lithics increased. During the 1800s, the three-age system – Stone, Bronze and Iron Ages – became widely used. In the mid 1800s, Sir John Evans taught himself how to flake stone, becoming the first to create replications of lithic technology (Kooyman, 2000). Refitting experiments were introduced in the late 1800s, with the early 20th century seeing an increase in experiments on flaking techniques. This is also when major experimental work was first published regularly. Use-wear analysis (Cotterell & Kamminga, 1987) and the mechanics of flaking were very popular study topics during the 1970s.

While such work is still continuing, the basis of current knowledge on fracture mechanics was accumulated during the 1970s through 1990s (Cotterell & Kamminga, 1987). Since the start of the study of lithics, time periods studied have ranged from Paleolithic Africa to modern-day Papua New Guinea.

The Acheulean

Based on the earliest evidence found in the Ethiopian Rift Valley, tool use began about 2.5 million years ago (Ambrose, 2001). *Homo habilis* and the later Australopithecines are recognized as the earliest tool users (Ambrose, 2001). This era of tool use is referred to as the Paleolithic, which is divided into the Early, Middle, and Late Paleolithic. The Early Paleolithic was characterized by the Oldowan and Acheulean tool types. An increase in retouched tools, the advent of Levallois tools, and the widespread controlled use of fire are all landmarks of the Middle Paleolithic, which was the time period of the earliest archaic *Homo sapiens*. During the Late Paleolithic we see the use of blade technology, bone, antler, and ivory use, and many examples of rich symbolic art (Toth & Schick, 2007).

JGD Clark defined the tool types from the Early Paleolithic as Mode 1 and Mode 2 tools, also known as Oldowan and Acheulean tools, respectively. Oldowan tools were used from 2.5 to 0.5mya, and their manufacture and use was at times temporally and spatially overlapping with that of Acheulean tools (Ambrose, 2001). There is a large amount of variability in Oldowan tools, arguably a reflection of the difficulties that differences in core morphology created for knappers (Ambrose, 2001). The creation and use of Acheulean tools began with *Homo erectus/ergaster* about 1.7 to 1.8 million years ago. The Acheulean spanned several species, including *Homo erectus/ergaster* and *H. heidelbergensis*, lasting until 250,000 years ago (Toth &

Schick, 2007). The Acheulean was a time of major climate change, with hominids extending into "cooler, more temperate climates" (Toth & Schick, 2007). The Acheulean Industrial Complex is characterized by large cutting tools, including cleavers, picks, blades, and handaxes (Ambrose, 2001). While the associated names have functional connotations, the groupings are based on tool shape. While the category names imply a difference in tool type, some of the differences in form between types may actually be a reflection of how finished a tool was at the time of discard. Instead of being completely separate tools with differing functions, some may simply be unfinished (Ambrose, 2001).

The Acheulean is typically characterized by large cutting tools including handaxes, picks, and cleavers. Acheulean handaxes typically have a teardrop shape and are between 10 and 17 cm in length (Ambrose, 2001). Handaxes are created through bifacial trimming and shaping, and based on the specificity of their shape, it is likely that their creators had a "well-defined concept of shape and proportion" (Ambrose, 2001). Towards the beginning of the Acheulean, the typical handaxe shape was quite rough and less symmetrical than the late Acheulean handaxes. Late Acheulean tools tend to be far more regular and symmetrical. Hodgson and McNabb propose that this difference is due to a difference in flaking technologies (2005). Handaxes are created through percussion flaking, which entails a dynamic loading of force. In this type of flaking, a hammer or percussor, which can be classified as soft or hard, is used to strike the object piece, which is held in hand or against the leg (Kooyman, 2000). Soft hammers include antler, bone and wood, while hard material hammers are often made of stone (Kooyman, 2000). Hodgson and McNabb believe that early handaxes may have been made using only hard hammer percussion, while later handaxes were likely finished using soft hammer percussion (Hodgson & McNabb, 2005). Stout et al. experimentally investigated the potential use of platform preparation at the

Boxgrove site as a technique for handaxe refinement, concluding that expert knappers successfully use platform preparation in cross-sectional thinning and that similar successful use was exhibited at Boxgrove (2014). Cross-sectional thinning is one of the most difficult to achieve characteristics of the refined Late Acheulean and cannot be done successfully by inexperienced knappers (Stout et al., 2014). Platform preparation is counterproductive if done incorrectly, making it a skill that could easily disappear, thereby potentially explaining some of the spatiotemporal differences seen in handaxe refinement (Stout et al., 2014). Overall, the change in Acheulean handaxe refinement is seen around 0.5 million years ago (Ambrose, 2001; Stout et al., 2014). In the Acheulean, distances of resource transport were greater than in the Oldowan, but rarely over 20km. The Acheulean in general is also notable for its "limited mobility and regional interaction" (Ambrose, 2001).

There is much debate over the uniformity of the Acheulean handaxe tradition over time and space. Often the Acheulean is described as "stagnant" in terms of changes in tool production. However, there is a high amount of regional variation within each type (Hopkinson, Nowell, & White, 2013). Cleavers tend to be widely found in Africa and India, but rarely in northwestern Europe (Hopkinson, Nowell, & White, 2013). In general, handaxes are widespread throughout the Old World, but not found east or north of the Movius Line, with some exceptions in China and Korea (Hodgson & McNabb, 2005). However, the greatest variability in handaxe form exists on the "local, assemblage, and short-term scales" (Hopkinson, Nowell, & White, 2013). According to recent studies, there is also evidence for the use of prepared core techniques, such as the Levallois, in the Acheulean, which had otherwise only been attributed to the Middle Paleolithic (Hopkinson, Nowell, & White, 2013). Because of the wide dispersal of Acheulean handaxes and the high level of regional variation, it is difficult to define what can or cannot be considered part of the Acheulean. Lycett & Gowlett used discriminant function analysis to compare handaxes from various sites, concluding that the same general “bauplan” was in use, meaning that knappers across varying regions created handaxes following the same approximate shape schema. Lycett & Gowlett were also able to identify regional differences at “extremely broad geographical levels” (2008). They suggest that function may also vary across region with the identified differences in shape (Lycett & Gowlett, 2008).

This study will be using refined Acheulean handaxes made of flint, similar to the archaeological assemblages from Boxgrove (Stout et al., 2014). The Acheulean Industrial Complex itself also includes picks and cleavers (Ambrose, 2001). In addition, categorization of a tool into one of the aforementioned Acheulean categories can be very difficult. Acheulean handaxes are only one type of biface, with a variety of other bifaces being found around the world. Prehistoric bifaces include Acheulean bifaces, Mousterian bifaces, leaf points, and North American bifaces. When studying bifaces, it is important to consider that resharpening and reuse, which are likely to have occurred frequently, can drastically affect a tool or debitage assemblage (Roe, 2002). The materials that are most commonly used in the creation of bifaces are homogenous and isotropic varieties of siliceous stone, including cherts. These are prime materials for flaking because their response to force is predictable (Cotterell & Kamminga, 1987). Frequently used materials include natural glasses, cherts, quartz, and quartzite. Cherts tend to have “greater fracture toughness” than natural glasses and quartz, and quartzite tends to have more flaws (Cotterell & Kamminga, 1987). Since Acheulean handaxes were created and used during a time period about which we often have limited sources of information, gathering as much useful information from lithics as possible is crucial. Many analytical approaches have been developed and applied to lithics.

Analytical Approaches

There are many analytical approaches used on debitage assemblages and many different measuring techniques as well. Every approach has different benefits and limitations (Morrow, 1997). Most analytical techniques work by identifying patterns in an assemblage of flakes (Shott, 1994). The approaches used in this study are experimental replication and the reduction sequence approach, discussed below.

Reduction Sequence

The concept of reduction sequences was introduced by Holmes in 1894 and is applied only to assemblages of stone tool debitage (Shott, 1994; Tostevin, 2011). Flakes that are removed during the creation of a stone tool can be viewed chronologically, in other words presented as a reduction sequence. Each piece in the reduction sequence is the result of a decision and action of the knapper. Since the concept's introduction, there has been discussion over whether reduction should be seen as a continuum or as a set of stages, and "explicit reduction models typically involved a sequence of discrete stages through which knapping proceeds" (Shott, 1994). Shott argues that each has an appropriate context and that "most analysts… intuitively recognize a reduction continuum even when they employ stage models as a matter of convenience" (1994). The stages that are used and defined vary widely by study (Shott, 1994). Several studies, discussed later, have attempted to create predictive models for reduction sequences using the known sequences from experimental replication assemblages. Using an equation to estimate an ordinal "flake number" or percent completion based on the dimensions of a flake could help in identifying what part of tool production was completed at a site, and thereby aid in assessing decisions made about resource transport and management.

Replication

Replication assemblages are useful in the identification of potential equations for percent tool completion or flake removal number. Replication is the controlled recreation of stone tools. These experimental sets can then be used for the comparison of archaeological sets, or, as Yerkes & Kardulias explain it, they can be used to "produce a "baseline" of lithic artifact variability" (1993). While creating a predictive model based on an actual archaeological assemblage would be preferable, currently the only way of identifying the process involved in reduction after the fact is through refitting. In refitting the location and orientation of flakes within the original core are determined by matching debitage pieces together. Given the difficulties involved in refitting and that it typically only has a success rate of 20%, which would likely not be sufficient to create a high quality reduction proportion equation, using experimental replication assemblages in which flake order is recorded is preferable (Laughlin & Kelly, 2010). Replication experiments have been very successful, with "virtually identical by-products [being] produced" when the archaeological data is used as an experimental control (Yerkes & Kardulias, 1993). The use of experimental replication sets as comparison for archaeological sets has frequently been criticized. Studies of replication assemblages have shown that equifinality is not as much of a problem as it was once thought to be, meaning that different methods *do not* lead to the same results (Shott, 1994). However, when all of the techniques used in the creation of an archaeological assemblage are not known or used in the replication assemblage, it is possible for systematic differences to arise. This was the case in a comparison of flakes created by expert modern knappers using platform preparation to a subset of the Boxgrove assemblage. Surprisingly, platforms were systematically smaller and thinner in the flakes of the archaeological assemblage. Stout et al. hypothesize this may have been due to an additional preparation technique of the dorsal release surface that was used by Boxgrove knappers, but not by the expert modern knappers (Stout et al., 2014). This shows that while replication can be useful and a good representation of past methods, slight differences in techniques can lead to significant differences in flake morphology. Replication studies are important for the understanding of debitage assemblages found at archaeological sites and of the behaviors involved in knapping. By numbering flakes as they are removed and logging the details of flake removal (e.g. hammer type, platform preparation, etc), replication studies can build the data set necessary to create a reliable and controlled reduction sequence model.

Biface Reduction Sequence Studies

As previously discussed, the goal of the accompanying study is to create a predictive model that can accurately place a flake within the larger context of its reduction sequence. Once such a model has been created using experimental replicates, its application to archaeological assemblages could provide valuable insights. By knowing how much of the tool creation process was completed at a specific site, inferences can potentially be drawn about resource management and transportation. While the concept seems straight forward, there are a variety of immediate problems researchers encounter when attempting to study reduction sequences, beginning with defining what parts of tool manufacture are included in the definition of "reduction." Reduction can also be approached either as a continuum or a set of stages, with enthusiastic proponents on both sides (Ensor & Roemer, 1989; Rozen & Sullivan, 1989). The studies that use a stage approach all define stages differently, which means that results are very difficult to compare across studies. In some cases, reduction is defined to include only retouch, after the manufacture of the original tool has long since been completed (Hiscock & Tabrett, 2010). These cases will not be discussed in detail here as they are more a measure of curation and tool maintenance than tool manufacture. The few studies that have attempted to identify a reliable index of completion for biface production based on metric measurements will be discussed below.

Stage Model

The earliest reduction sequence studies approached the process of reduction as a set of stages. Collins delineated reduction as five steps, "acquisition of raw material, core preparation and initial reduction, optional primary trimming, optional secondary trimming and shaping, and optional maintenance/modification" (Collins, 1975). Many different definitions of the stages of reduction exist, another example being "(1) obtaining the blank, (2) initial edging, (3) primary thinning, and (4) secondary thinning" (Patterson, 1990). The traditional typological approach, which splits flakes into primary, secondary, and tertiary groups based on cortex amount, was used to split flakes into reduction stages (Bradbury & Carr, 1995). However, in longer reduction sequences, cortex cover only exists on a portion of flakes, making it a poor choice (Bradbury & Carr, 1995). Very often the overall process of tool creation will be split into "core reduction" and "biface reduction," which Magne and Pokotylo argue have "substantially different" technical aspects (1981). The original purpose of the identification of reduction sequence stages was to gain an understanding of a tool's "manufacturing trajectory." This term refers to the locations where the various reduction stages were completed. According to Patterson, there are several common tool manufacturing trajectories: (1) tool manufacture is started and finished at the site of the resource; (2) the flake blank is created at the site of the resource and is transported to another site, where bifacial reduction occurs; (3) preforms are taken to a location away from the resource site and all reduction occurs there (Patterson, 1990). Flake size distribution has not been found helpful in sorting flakes into reduction stages (Patterson, 1990).

Magne and Pokotylo performed a pilot study on reduction stages in an attempt to minimize the number of variables needed for analysis (1981). Upon reviewing many other studies, they decided to measure five continuous and three ordinal variables (weight, length, width, platform width, dorsal angle, dorsal and striking platform scar counts, cortex cover in 25% intervals). Applying hierarchical clustering, metric multidimensional scaling, and multiple discriminant analysis to these variables they were able to define six debitage categories: core reduction, middle reduction, late blank reduction, biface reduction, early shatter, and late shatter (1981). They then tested this classification by applying it to archaeological sets from Upper Hat Creek Valley. These sets had previously been categorized using a different technique. Pokotylo and Magne's test of the new classification system "was quite successful in showing that the classification is a reliable, efficient means of inferring lithic reduction activities" (Pokotylo & Magne, 1981). Here again several issues with the stage names arise, the first problem being that another group could have arrived at the same categories, but named them something completely different, making comparison across studies difficult. In addition, the names of the debitage categories in this, and many other studies, are based on presumed stages in technique. The naming of these stages implies that certain cognitive shifts occurred between debitage categories, which, while possible, is not supported solely by the existence of such categories. It is also possible that such identified stages could correspond more closely with a confounding variable, hammer type for example, rather than with the actual reduction sequence. In this case, identifying such groups as “reduction” groups could result in incorrect conclusions when applied to assemblages in which hammer use differed from that in the experimental assemblage.

Amick and Mauldin provide another example of stage based reduction sequence analysis. Amick and Mauldin studied the bifacial core reduction of one Georgetown chert nodule (1988). They measured six discrete and seven continuous variables. They used discriminant function analysis to identify stages and found that all but one of their categories were statistically significant. A function was created to sort flakes into each class and discriminant analysis was used to assess the strength of their predictions using the recorded variables. The overall goal was to achieve the highest possible rate of correct classification using the lowest number of variables. They were able to correctly classify platform remnant-bearing flakes in 76.2% of all cases, which is significantly better than chance. The reduction of the chert nodule did involve use of multiple hammer types, the effect of which was considered during analysis. While it did serve as a suitable pilot study, which was indeed the intention, investigating more material types and using more than one reduction assemblage per material type would be preferable.

Continuum Model

There has been much discussion between proponents of the continuum model and the stage model (Rozen & Sullivan, 1989; Ensor & Roemer, 1989). Those who work with the continuum model frequently argue that stage models tend to impose arbitrary categories onto the reduction process. By using names such as “biface thinning flakes” they are also tying functions to classes of debitage without direct evidence that the function is related to the class (Rozen & Sullivan, 1989). Many, however, also believe that both methods are useful and can complement each other well (Bradbury & Carr, 1999). There have been relatively few studies that attempt to create a function for percent completion or flake order. The studies of Bradbury & Carr (1999); Ingbar, Larson & Bradley (1989); and Shott (1996) are all examples of such studies.

After finding accuracy problems with Magne and Pokotylo's approach (1981), Ingbar et al. analyzed the reduction sequences of four bifaces created from materials from Wyoming using the continuum approach (1989). Using the variables of log of flake thickness (LOGTHK), log of dorsal scar density (LOGDSD), platform scar density, maximum flake width, and log of flake area (LOGAREA) they created five predictive models for flakes' removal numbers. Their simplest model that "achieved adequate r-squared values and a slope coefficient between 1 and 2" was:

$$\text{Predicted } y = -63.75(\text{LOGTHK}) + 18.24(\text{LOGDSD}) + 29.62(\text{LOGAREA})$$

They then applied this equation to an archaeological assemblage from the Early Plains Archaic levels.

Shott's continuum model experiment used a replicated fluted biface (Shott, 1996). Shott also attempted to identify reduction stages, but unlike previous stage model analyses, was unable to. Cluster analysis was used to analyze the metric data and Shott found that "clusters do not sort out along the continuum and do not represent discrete stages of it." Similarly to Ingbar et al.'s study, Shott identified multiple predictive models for removal number, using flake weight, scar density, and platform width. Given that Shott used a fluted biface, his predictive models may, for example, not be applicable to non-fluted bifaces. The fluted biface of this study was also created using only soft-hammer percussion, making its applicability to an archaeological set debatable. Shott identified the following two removal number equations:

1: removal number= 10.0 scar count-15.3 log -weight+17.0 log-platform width
2: removal number = 12.1 scar count - 15.5 log-weight + 4.9 platform width

Model 1 had an adjusted r^2 value of 0.78 and model 2 had an adjusted r^2 value of 0.82. When the predicted values of model 1 were plotted against the actual removal numbers, the slope coefficient was 0.62. The slope coefficient for model 2 was 0.54. Ideally, the slope relating the predicted values to the actual values would be 1.0, suggesting that these models may not be very useful.

Using flintknapping replication experiments, Bradbury & Carr analyzed several reduction assemblages using both continuum and stage models. In their continuum model analysis, they applied the formulae provided by Ingbar et al. (1989) and Shott (1996) (Bradbury & Carr, 1999). They found that both provide a good "general characterization" of the assemblages. Their own final analysis using the continuum model entailed developing a model for percent completion:

> Percent complete = (0.0898 * facets) + (0.0713 * log maximum width) + (0.01638 * log scargrams).

In this model, "scargrams" is calculated by dividing the dorsal scar count by the flake weight. The r-squared value for the percent complete model was 0.86, which suggests that the model explains much of the observed variance. However, the r-squared value is skewed here because the regression was forced through the origin (Casella, 1983). When predicted values were compared to actual values, they found a slope of 0.86, suggesting that the predicted values will typically be lower than the actual reduction values. As can be seen, the variables important to the model are the number of facets, the maximum width, the weight, and the number of scars. Bradbury and Carr encourage the use of percent complete instead of event number, because of its practicality as a standardized scale. All of their reduction assemblages were created using chert from the same source, thereby eliminating potential difficulties caused by material type. They also identified a way to separate core reduction flakes from tool production flakes using discriminant function analysis. The percent complete model would then only be applied to the tool production flakes. While Bradbury and Carr's experiment is likely the most thorough of the discussed studies, it is possible that one could create a more reliable percent completion model by focusing on just one specific tool type, rather than incorporating a range of tool types. Certain assemblages used in the experiment were also created using solely one type of percussion. While the entire experiment at large contained both hard and soft hammer use, a better predictive model may have been achievable had each reduction contained more natural hammer choice and been of the same reduction type (bifacial versus unifacial).

Recently, a study by Shipton and Clarkson identified core scar count divided by surface area of the remaining core as a reliable predictor for the amount of reduction or retouch the core had undergone (2015). They found that scar count/surface area increased as the percent of the original mass that still remained decreased (Shipton & Clarkson, 2015). While this study is investigating scar count on the remaining tool, not the flakes, the same principle should apply to the dorsal scar count on flakes, given that flakes are a reflection of part of the core prior to the most recent hammer strike.

Overall, the continuum approach seems promising, while using the stage model is also still valuable. Both can be applied to the same assemblage and can provide valuable information. The measurements found most useful in both analyses were weight, length, width, platform width, scar count, dorsal angle, cortex cover, and platform faceting. For the purposes of learning about resource transportation, a continuum model would be the preferable approach given the high variability in the kinds and characteristics of stages defined using the stage model approach. The use of a continuum model instead of a stage model would also eliminate the need for separation of flakes into arbitrarily named categories such as “core reduction” and “tool production” and could potentially better buffer against the effect of hammer type.

Existing and Potential Limitations of Reduction Studies

As is clearly seen in the examples discussed above, the existing reduction sequence studies face several easily identifiable problems. In the stage model studies, the stages are often inferred to correspond to intentional stages recognized by the knapper. The stages identified may also correspond more closely to a confounding variable such as hammer type rather than actual stages of reduction. Additionally, each research group uses different measuring styles, often making comparison and application of reduction equations difficult.

Beyond the lack of regularity and clear definitions, the creation and application of a reliable index for level of completion also faces many other potential limitations. Possible confounding variables include core material, hammer type, knapper handedness, knapper skill-level, and knapper “style”. However, the application of a completion index faces a larger, more general problem, namely the high amount of variability in archaeological sets. Unless a finished tool is present, it can be very difficult to identify what type of flaking was used to create a specific debitage set, especially if only debitage from a part of the full reduction is present. Furthermore, assemblages found at archaeological sites often consist of debitage from more than one tool reduction. These factors, combined with the lack of standardization in measurements, definitions, and analytical approaches, make reduction sequences a difficult research subject.

Lack of Repeatable Measurements

Unfortunately, there are many differing measuring methods and measurement errors can occur easily in lithic analysis (Shott, 1994). Many differing definitions of flake length, width, thickness, and other measurements exist in the literature, which can make it very difficult, if not impossible, to compare data across studies. Length and percent of dorsal cortex cover have been shown to be unreliable measurements in several studies, and though controversy exists, platform angle may also fall into this group (Shott, 1994). Proper application of potential reduction sequence models requires utmost clarity in measurement definitions. Ideally, the use of difficult to replicate measures would be avoided.

Core Material

Cores have the potential to be extremely variable between reductions. Cores can vary in stone type, size, quality, and uniformity, among other attributes. Raw material type has been shown to affect patterns of flake breakage (Amick & Mauldin, 1997). Sub-par materials and imperfections in the core can also easily affect the reduction process (Amick & Mauldin, 1997). Amick and Mauldin studied the effects of raw material type, hard versus soft hammer percussors, and reduction type (core reduction versus tool production) on flake breakage type frequencies. They created twelve bifacial tools, categorizing the debitage into the categories of complete flakes, proximal flakes, split flakes, medial-distal fragments, and non-orientable fragments. Obsidian, basalt, quartzite, and several cherts and chalcedonies were used for replications. Amick and Mauldin found that reduction type (core reduction versus tool production) does not affect flake breakage type frequencies, but did find that in both the core reduction phase and tool production phase, differences in raw materials resulted in statistically significant differences in flake breakage types. Overall, differences were found to be trend level when reduction type was not taken into account. In general, a greater tendency for split flakes existed in basalt/quartzite than in the other tested materials (Amick & Mauldin, 1997). Given that core material has been shown to affect flake and tool morphology, it should always be controlled for and taken into account when analyzing archaeological assemblages.

Goodman also found that different stone materials react differently to abrasion (1944). Materials have varying levels of hardness and toughness, the second of which contributes to determining when a material will fracture, thereby affecting stone tool reduction (Goodman, 1944). Peter Jones also found that raw material type affects a core’s potential usages (1979). Different raw materials can and cannot effectively be resharpened after manufacture, which could drastically affect resource management decisions. He also showed that different raw materials result in longer or shorter manufacture times, again an important factor for decisions on material use. Raw material type and how these varying materials flake also influences the final tool shape (Jones, 1979). Lastly, the raw material of an assemblage being analyzed has been shown to significantly impact inter-analyst variability and accuracy, which places even more importance on using well defined and easily replicable measurements for reduction sequence studies (Proffitt & de la Torre, 2014). It is possible that a reduction completion model created using one material type may not be applicable to an assemblage of another type. Ideally, separate predictive equations could be created for the most frequently used materials, eliminating raw material as a potential confound and helping us better understand the risks of inter-analyst variability. Unfortunately, creating reliable and effective predictive equations for each material would be quite laborious. Since raw materials have varying reactions to abrasion, variation in material could also affect how well the flakes preserve and how comparable they are to freshly made replicates.

Hammer Use

Many studies have shown differences in attributes between hard and soft hammer flakes, but few studies agree on what those differences are (Pelcin, 1997). Several of these studies even directly contradict each other regarding what attributes are affected and how (Pelcin, 1997). While Pelcin’s study strictly controls for many factors, including platform thickness, he remarks upon the fact that these controls may, from one perspective, be reducing the quality of the study. Knappers account for what percussor type they are using when deciding where to strike the objective piece, meaning that a direct comparison of perfectly controlled hard hammer versus soft hammer flakes may not be as helpful as hoped (Pelcin, 1997). Pelcin found that antler indentors resulted in longer flakes relative to the steel ball bearing indentor used as a hard hammer. Flake length, flake thickness, bulb length, bulb thickness, and expansion angle all showed differences between indentor types (Pelcin, 1997). This study shows that differences in flake attributes exist between hard and soft hammers because of both hammer type and other factors that co-vary with hammer type. However, as Magnani et al. discussed, a combination of various factors, such as location of force application, angle of blow, and hammer hardness, can result in similar flake features. While it is possible to examine the effect of changes in one variable at a time, investigating the effect of combinations of factors is far more difficult (Magnani et al., 2014). Because of this, identifying what variable was actually mainly responsible for a feature in an archaeological assemblage is quite difficult (Magnani et al., 2014). It was also established that multiple techniques can result in the same features; for example a soft hammer on a core edge has the same effect on flake elongation as the use of a hard hammer on a platform edge (Magnani et al., 2014). After material type has been accounted for, one could hope that hammer type would be a smaller problem because experienced knappers may be likely to choose the same hammer type as other knappers throughout the reduction process. In analyzing data, it would be important to remain aware of how hammer choice changes throughout reduction (e.g. primarily soft hammer for finishing) and how hammer type alone affects flake morphology.

Handedness

Toth was the first to publish an attempt at identifying a knapper's handedness in 1982. He predicted that cortex placement could be used to reliably assess knapper handedness because rotation of the core during reduction would create an identifiable pattern (Toth, 1982). However, since then this method has received much criticism (Uomini, 2009). More recently, Rugg and Mullane attempted to identify handedness using a measure of skew in the cone of percussion. Using their method, they were able to correctly sort assemblages 75% of the time (Rugg & Mullane, 2001). Others, however, have criticized both of these approaches for their attempt to find a single diagnostic feature. Unsatisfied with this approach, Bargallo and Mosquera assessed multiple features, and while they did not identify a single diagnostic feature, they do suggest that a combination of variables may provide a reliable assessment of handedness (Bargallo & Mosquera, 2014). While there is some disagreement regarding to what level handedness affects features and which features are affected, there does seem to be a general consensus that knapper handedness does have an effect, making it a potential confound for level of reduction completion models. However, it also seems that the majority of flake traits affected by handedness are related to flake symmetry or skew. By avoiding the use of flake attributes affected by flake skew, the effect of knapper handedness would hopefully be minimized.

Knapper Experience Level and Style

The quality and shape of a product will differ between individual manufacturers for a variety of reasons, including personal style and experience level. To be able to shape a desired tool, a knapper has to be able to control the shape of the flake he is removing. Nonaka et al. compared how well expert, intermediate, and novice knappers were able to remove the flake they predicted. Novices tended to predict the removal of relatively small flakes when compared to experts and intermediates. Novices also removed significantly smaller flakes than intermediates and experts (Nonaka et al., 2010). Experts were better able to predict their flake sizes and were able to successfully remove longer, more difficult flakes. Experts also showed selectivity in their choice of flaking surface (Nonaka et al., 2010). Winton also investigated the differences between experienced and novice knappers, finding that handaxes created by less experienced knappers tend to be relatively shorter and thicker, and asymmetrical (Winton, 2005). Similarly, significant differences in the dimensions of flakes produced by novice and experienced knappers have also been found in adzes from Irian Jaya (Stout, 2002). Stout suggests that a leading reason for differences in flake dimensions may be “the greater ability of experts to exploit and manipulate core morphology” (Stout, 2002). It has also been shown that platform preparation, crucial for bifacial thinning, is a difficult technique to master and is counterproductive when usage is attempted by a novice knapper (Stout et al., 2014). The same study also identified a systematic difference in flake thickness between novice and expert knappers (Stout et al., 2014). As here discussed, multiple studies have shown differences in flake dimensions correlated with knapper skill level, suggesting that skill level must also be controlled for in reduction sequence analyses. Additionally, Williams and Andrefsky identified significant differences in flake attributes of assemblages created by five different experienced flint knappers, suggesting that personal style influences debitage variability (Williams & Andrefsky, 2011). Maximum linear dimension, maximum linear width, maximum platform dimension, and weight varied significantly between knappers in early stage bifaces, implying that these variables would not be well suited for use in a reduction model (Williams & Andrefsky, 2011). As previously mentioned, the potential for differences in techniques applied to replication assemblages and archaeological assemblages can also limit the applicability of a reduction sequence model (Stout et al., 2014). Knapper skill level and style would unfortunately be two of the most difficult obstacles in the application of a reduction sequence equation to an archaeological assemblage.

Identification of Sets

An archaeological assemblage could contain any number of debitage sets from separate reductions, all mixed together into one assemblage. To make things even more difficult, some of those debitage sets could be from non-bifacial reductions or from incomplete reductions. It has been shown that the flake size distribution of complete bifacial reduction assemblages creates an exponential curve (Patterson, 1990). In an archaeological set, some slight deviation from this pattern is normal (Patterson, 1990). According to Patterson, this is the only type of reduction that results in this type of curve, which could potentially be useful in differentiating types of assemblages. However, a mixture of several bifacial reduction sets will have this same distribution pattern, making it useless in differentiating bifacial reduction assemblages from each other. Unfortunately, the separation of assemblages is very difficult without other clues such as material type and is a continuous problem in the recovery and analysis of archaeological debitage sets. If the same portion of reduction was continuously completed at a site, a percent completion model would still be useful. The problems of mixed assemblages and ranging knapper skill level are the most difficult to address in the creation and application of a percent completion or reduction sequence equation.

Clearly a wide range of potential difficulties exists in reduction studies. Additionally, “hammer shape, location of force application, angle of blow, [and] hammer displacement speed” have been shown to affect flake morphology, even when the effects of a single variable may not be clearly defined or easily identifiable (Magnani et al., 2014). Magnani and his colleagues stress the importance of interactions between all of the variables and the resulting difficulty in predicting flake morphology (2014). Ideally, reduction assemblages used for reduction sequence studies would be as realistic as possible, appropriately reflecting such variation.

Identifying the Gap

Based on the discussed limitations complicating reduction sequence studies, we can formulate what an ideal approach to a reduction sequence study for transportation analysis might look like. All measurements used in the model would be measurements that are minimally affected by skew and symmetry, thereby avoiding effects of handedness. The study would also have to establish clear and easy to apply definitions of all measurements, and avoid difficult to replicate measurements, thereby optimizing the applicability of the created model. The resulting model would be created specifically for one material type, without assuming that it is simultaneously applicable to other material types. Given the influence of hammer type on flake morphology, hammer type would additionally have to be controlled for. Unfortunately, knapper experience level and mixed assemblages pose continued problems that are more difficult to solve.

With the creation of reliable percent completion models for varying material types, one of the many limitations facing lithic assemblage assessment would be reduced. Tracing resource transportation across a landscape would be possible with a percent completion model and would give us insight into resource management practices during the Acheulean. That information in turn could begin to help us understand more about the Acheulean knappers' decision making and analytical skills. Outside of resource transportation, an accurate reduction sequence model would additionally open doors for other types of analysis that require relative flake order. This study attempts to create a proportion of reduction completion model for refined Acheulean handaxe technology made of flint. This would then be applied to data from the Boxgrove site in West Sussex, England (Stout et al., 2014; and unpublished data). Ideally, it would be possible to identify what portion of reduction occurred at the site and thereby shed more light on the processes involved at and usage of the site.

Materials and Methods

Three Acheulean handaxe replication sets were used in this study. Two of the assemblages were created by knapper Dr. Dietrich Stout and one by Dr. Bruce Bradley. Both of these knappers can be classified as expert knappers. The handaxes produced by both are comparable in refinement to those in the Boxgrove assemblage, though slightly larger (Table 1; Stout et al., 2014). Platform preparation was identified at Boxgrove and the proportion of flakes with platform preparation is very close to the mean of the proportions from the assemblages created by the expert experimental knappers (Stout et al., 2014). The experimental assemblages were created using a combination of soft and hard hammers. The knapper chose hammer type individually for each blow, depending on the challenge presented by the core. In this way, hammer choice was as naturalistic as possible, ideally making it more applicable to archaeological assemblages. However, it is also possible that this creates an additional confound, given that the experimental knappers had modern decision making skills. There is evidence that both hard and soft hammers were used at Boxgrove. Antler pieces and bones containing fragments of flint were found at the site, suggesting they were used in knapping (Stout et al., 2014). While the material used to create the experimental handaxes was flint, it was purchased from the Cardy of Ingham quarry in Suffolk, England (Stout et al., 2014). The material used to create the Boxgrove handaxes has been concluded to be from the site of Boxgrove itself (Roberts & Parfitt, 1999).

During the experimental replication, each flake was numbered in order of removal from the core during tool production. Presence/absence of platform preparation or trimming and hammer type choice were also recorded for each strike. A proportion of relative removal was calculated by dividing the removal number by the total number of removals. In essence, it is a proportion representing reduction completion. Two flakes could have the same removal number or relative removal value if they were removed during the same hammer strike. Each assemblage contains the flakes produced from the start of core reduction through the completion of a refined Acheulean handaxe. Only whole flakes were included in this study, whole flakes being defined as flakes with minimal, <15% estimated, damage from which a measure of length and width could accurately be taken. Only flakes larger than 20mm in their largest dimension were considered (Stout et al., 2014). A summary of information on the three assemblages used in the study can be found in Table 1. Figure 1 is a photograph of one of the complete experimental assemblages aligned chronologically, with those flakes in the left column being from the beginning of reduction and those in the rightmost column being from the end of reduction. Clearly, small flakes are spread throughout the entire reduction sequence, meaning that size alone is not a good indicator of reduction progression.

Figure 1: One of the experimental assemblages organized with the first flakes of reduction on the bottom left and reduction progressing through the columns and towards the right.

Table 1: A summary of basic information on each experimental assemblage used in the study

Assemblage	Knapper	Core Material	Final handaxe length (mm)	F.H. width (mm)	F.H. mass (g)	Number of whole flakes	Percussion Types
Set 1	DS	Flint	211	115	705	121	Hard and soft
Set 2	DS	Flint	224	114	652	95	H. and S.
Set 3	BB	Flint	155	85	313	68	H. and S.
Boxgrove (n = 18)	-	Flint	Mean = 134 mm	M. = 82	M. = 375	491	H. and S.

(Stout et al., 2014)

Debitage Measurement

Information on seventeen variables was collected for each flake. One of these measurements (cortex) was categorical, while all of the others were continuous. In an attempt to normalize the data, the majority of the following measurements and computed variables were later log transformed. All distance measurements were recorded in millimeters. Measurements were not recorded when relevant flake landmarks could not be identified confidently.

Dorsal scar count – Dorsal scar count was determined via visual inspection of the flake. Only scars larger than 10mm were included in the count. A scar was identified as an area enclosed by noticeable ridges that clearly separated the scar from the neighboring flake area.

Cortex – The existence of cortex on the flake was recorded as either “yes” or “no.”

Length – Flake length was defined as the length between the point of percussion and the most distal point along edge of the flake. (Debenath & Dibble, 1994)

Maximum Width - Maximum flake width was defined as the maximum width measured perpendicularly to the axis defined by length.

Width measurements – Three additional width measurements were taken. Measurements were again taken perpendicularly to the length axis and measured at ¼ the full length, ½ the full length, and ¾ the full length.

Thickness measurements – Three thickness measurements were taken, one each at the ¼, ½, and ¾ length positions. Thickness was defined as the thickest section along the axis defined by the accompanying width measurement.

Mass – Mass was determined using an electronic scale and recorded in tenths of grams.

Platform breadth - Platform breadth was defined by the intersection of the platform with the lateral margins of the flake.

Platform thickness – Platform thickness was defined as the longest distance from the ventral to the dorsal edge of the platform, measured perpendicularly to the axis of platform breadth.

Exterior platform angle – Exterior platform angle was measured using a goniometer. The "exterior" was defined as the angle created by the platform and the dorsal flake surface.

Interior platform angle – Interior platform angle was also measured with a goniometer and was defined as the angle created by the platform and the ventral flake surface.

Bulb length – Bulb length was measured from the point of percussion to the visually determined point of inflection.

Bulb thickness – The measure of bulb thickness represents the thickest part of the identified bulb, measured orthogonally to the plane defined by flake length and width.

Computed Variables

Area – Area was calculated by multiplying one third of the length measurement by the sum of the three width measurements (*not* including the maximum width measurement).

Scar count/area – This variable was computed by dividing the dorsal scar count number by the flake's calculated area.

Average thickness – Average thickness was calculated by taking the average of the three thickness measurements.

Average width – Average width was calculated by taking the average of the three additional width measurements (*not* including the maximum width measurement).

Average deviation of thickness measurement from average thickness – This measure was created by taking the average of the absolute values of each thickness measurement independently subtracted from the average thickness. This measure was intended to represent the variability of width along the length of a flake.

Average deviation of width measurement from average width – This was calculated using same method as the previous variable, using the width measurements instead of thickness measurements.

Platform Area – Platform area was calculated by multiplying platform breadth by platform thickness.

Computations Using the Geometric Mean

The geometric mean constant was calculated using flake length, the three additional flake width measurements, the three thickness measurements, bulb length, bulb thickness, and platform thickness and breadth (all linear measurements except maximum width were included, since width was already captured by the use of the additional width measurement). The geometric mean was identified to control for overall flake shape and size. All of the included linear measurements were then divided by the geometric mean to control for overall flake size.

Residuals

Regressions relating log(mass) to log(platform area), log(maximum width) to log(length), and log(length) to log(average thickness) were completed. Standardized residuals of the aforementioned regressions were saved and their relationship to reduction completion assessed. If residuals are used in the reduction sequence model, one could obtain this variable in an archaeological application by performing the same linear regression on the archaeological assemblage and saving the residuals. Standardized residuals would be better for use in the model because they would control for an effect of size, while unstandardized residuals would not necessarily do so.

Principal Component Analysis

A principal component analysis was run on the measures divided by the geometric mean. Two principal components were identified, but once used in a multiple linear regression with other measures they were not as valuable in predicting reduction completion as the single variables alone.

Predictions

Overall Flake Attributes

At the beginning of reduction, it would make sense for a knapper to create larger flakes that quickly diminish the overall size of a nodule and to achieve roughly the correct handaxe shape. The shape of such flakes would not necessarily need to be finely controlled. However, as reduction progresses, it would be beneficial for a knapper to create smaller flakes with more controlled shape so that they can refine the handaxe shape (Winton, 2005). One might expect such flakes to be smaller in terms of most linear measurements, weigh less, and to be more consistent in width and thickness. It was hypothesized that variability in thickness and width along the length of the flake would decrease as reduction proceeded given that changes in the tool shape with each strike would ideally be small, controlled, and uniform later during reduction. Bulb thickness was hypothesized to decrease as reduction proceeded because a smaller bulb would leave a smaller, more uniform scar on the tool being created. Given that flakes, and thereby flake scars, are expected to be smaller later during reduction, it would also make sense for the number of dorsal flake scars per flake to increase (Shipton & Clarkson, 2015). When controlled for overall flake size using the geometric mean, it is predicted that several measurements, including thickness, will still decrease over reduction.

Variables Identified in Other Studies

Past studies have identified flake thickness, dorsal scar count and density, maximum flake width, flake mass, and platform width as potential reduction sequence predictors (see introduction). Flake thickness, maximum flake width, flake mass, and platform width are all likely to decrease as reduction progresses given that one can expect flake size to decrease. As previously discussed, the changes in dorsal scar count are also to be expected.

Statistics

Initial overview

All statistical analysis and computation of variables was completed using IBM SPSS Statistics. Scatter plots of the relationship between relative removal and potential predictor variables were created to visually assess potential trends. When the relationship appeared to be more logarithmic than linear, the variable in question was log transformed. Linear regression was used to evaluate the potential strength of single predictor variables. All variables that achieved an R^2-value of more than 0.075 were then used in a multiple linear regression. A cut-off of 0.075 was chosen because the R^2-values for most variables were either in this range or noticeably below it.

Linear Regression

Using the potential variables identified, a stepwise multiple linear regression was run. The stepwise method was selected because it adds and removes variables based on the p-value of F, removing variables later on if their p-value becomes too large. This would be preferable over the entry or remove methods which test all variables simultaneously, thereby not as thoroughly identifying which variables should or should not be included in the model. The stepwise is preferable over the forward or backward methods, which either only remove or only add variables consecutively, instead of identifying the best combination regardless of initial order. Multiple regressions were run both with and without the option of forcing the model through zero.

The four best models, identified based on standard error of the estimate values, were then identified and recorded. Predicted relative removal values were saved for each model and were regressed against the actual relative removal values. Ideally, such a regression would result in a slope of 1 and a very small standard error of the estimate (Shott, 1996; Bradbury & Carr, 1999). The best model was then selected based on those criteria. If there was no clear “best” model, the two best were chosen.

Identification of Predictive Equation

The predictive equation can be identified based on the beta and intercept values provided by the linear regression analysis. For the creation of this model, unstandardized rather than standardized coefficients were used because standardized beta values are rescaled so that the y-intercept is equal to zero, which should not be done unless there is very definite reasoning regarding why the variables used necessitate an intercept of zero (Casella, 1983). Forcing a model through the origin can skew R^2 and F statistics, making it difficult to compare models to each other (Casella, 1983).

(predicted variable) = intercept + $beta_1$ (independent $variable_1$) + $beta_2$ (independent $variable_2$) + … $beta_n$ (independent $variable_n$)

Assessment and Application

The potential value of the model was first assessed using the significance value of the model and the standard error of the estimate. The standard error of the estimate, provided in the SPSS output of the multiple linear regression, represents how well the model predicts the value. The smaller standard error of the estimate, the more reliably the model tends to predict the relative removal position of a flake. A histogram of the predicted relative removal values was created and compared to a histogram of the actual relative removal values to visually assess whether the model was over-predicting a certain removal range.

Next, because of noise in the data, data were binned to create two additional ordinal variables, one sorting flakes into the first, second, third, or last quarter of reduction, the other sorting flakes into the first or second halves of reduction. An independent samples t-test and one-way ANOVA were performed for the halves groups and quarters groups respectively to identify if the mean predicted reduction value was significantly different between portions of reduction. If the means differed significantly, the next step of analysis was performed. An ANOVA and independent samples t-test were also performed on the means of the actual relative removal values as well. The means of the actual relative removal groups were identified so that the means of the predicted values could be compared to the actual values.

If the means of the predicted relative removal differed significantly between the two halves and between each of the quarters, the next step of analysis was completed. Using a one sample t-test, the mean of the actual relative reduction values and of the predicted values were compared to the expected mean of 0.5. The means of the predicted values for each half and each quarter were then compared to means of the actual values (which did not differ significantly for the expected means) via one sample t-tests for each half/quarter. This was done to identify whether the means of the actual values and predicted values were approximately equal in each half/quarter, which would mean that the model could potentially be used to identify what quarter or half of reduction is represented at an archaeological site.

Once a strong potential model had been identified, it was applied to data from the Acheulean flint assemblage from Boxgrove (Stout et al., 2014 and unpublished data). The goal of application was not to predict each individual flake's position within the reduction sequence, but rather to characterize the assemblage as a whole. Because variables identified for the models required calculations using the three breadth and thickness measurements, the best model was assessed a second time recalculating variables using only the average thickness and width measurements. If beta coefficients were to change significantly, the coefficients of the model would be changed for the application to the Boxgrove data set. Once predictive values were calculated for the Boxgrove data set, the mean was compared using one sample t-tests to the means of the predictive values established for each quarter and half of the experimental sets. In this way, placing the predicted mean of the Boxgrove assemblage within the reference means was attempted.

Testing Assumptions for Linear Regression

Lastly, validity of the model created via multiple linear regression was assessed by determining whether all assumptions for linear regression were met. For assumptions to be met, all variables must be linearly related to the predicted variable, distribution of residuals must be normal, error variance must be constant, errors in one variable must be independent of those in another, and the model must include only relevant variables. The use of the stepwise linear regression ensures the inclusion of only relevant variables, given that predictors are removed from the model if a better encompassing variable is added. Linearity was visually assessed using scatter plots containing the variables included in the model. Normality of residuals was tested using the Shapiro-Wilk test for normality. Homogeneity of variance was determined by visually assessing graphs of the residuals plotted against the predicted values. An even spread of residual values over the span of predicted values means that variance is relatively homogenous. Independence was assessed by creating scatter plots of the residuals versus each of the independent variables. If the variables are independent, there should be no correlation between the residuals and the independent variables. It is also important to determine whether the independent variables are collinear, which can be determined by running collinearity diagnostics with the linear regression.

Results

Initial Overview

Creating scatter plots of relative removal versus potential predictor variables or their log transformed versions was helpful in identifying potential trends. Initially, many variables seemed to show a general trend associated with reduction. See Appendix A for scatter plots and R^2-values of the relationships between potential variables and the proportion of reduction completion. The variables with R^2-values over 0.075 were length/GM, bulb thickness/GM, average thickness/GM, average width/GM, log(length/GM), log(bulb thickness/GM), residuals from a regression relating log(length) and log(average thickness), log(mass), log(scar count/flake area), average difference in thickness across the length of the flake, log(platform area), and geometric mean. Variables identified in previous studies, such as log(maximum width), flake mass, and platform width, were also investigated

Multiple Linear Regression

Once viable candidate variables had been identified, several stepwise multiple linear regressions were run including various combinations of the previously identified variables. When all identified variables were included, the best model produced included log(scar count/area), log(maximum width), and the standardized residuals. However, in combinations that included everything but maximum width, models were created with higher R^2-values and lower standard error of the estimate values, suggesting they are better models. The four best models identified throughout all regressions, with and without log(width) and with and without forcing the regression through zero, were the following.

Model A:

2.052 + 0.579*log(scar count/area) + 0.279*log(mass) – 1.172*(average thickness/geometric mean) – 0.54*log(bulb thickness/geometric mean)

Model B:

2.013 + 0.574*log(scar count/area) – 0.013*standardized residuals + 0.278*log(mass) – 1.086*(average thickness/geometric mean) – 0.502*log(bulb thickness/geometric mean)

Model C:

0.855*log(length/geometric mean) + 0.257*(average width/geometric mean) + 0.495*log(scar count/area) + 0.365*log(platform area)

Model D:

0.731 + 0.544*log(scar count/area) + 0.763*log(maximum width) – 0.091*standardized residuals

Comparison of Models

As described in the methods section, the models were compared via regressions of the predicted values versus the actual relative removal values. Model A and B were identified as the best models with beta values closest to 1.000, the highest adjusted R^2 values, and the lowest standard error of the estimate values (Table 2a). Scatter plots of the predicted values for each model versus the actual relative removal proportion helped visualize the predictive abilities of each model.

Table 2a: Comparison of the various models identifies model A and B as the best potential models.

Model	**Beta coefficient**	**Adjusted R^2**	**Standard error of estimate**
A	1.000	0.434	0.205
B	1.000	0.435	0.205
C	0.997	0.428	0.207
D	0.969	0.338	0.230

Additionally, the contributions of each term to the adjusted R^2 value for both models were noted (Table 2b). While both models contain almost the same variables and have the same adjusted total R^2 value, variables contribute differently depending on whether or not they were included in the model prior to or after the standardized residuals of the log(average thickness) log(length) regression. The largest single contributor is without doubt log(scar count/area), with log(mass) also being a large contributor in both models. When the standardized residuals are involved in the equation, the contributions of log(bulb thickness/GM) and average thickness/GM are affected most severely.

Table 2b: The impact of adding each variable upon the R^2 value for each model.

Model A	Log(scar count/area)	Log(BlTh/GM)	Log(Mass)	AvgTh/GM	-	-
R^2 values	0.211	+0.07	+0.057	+0.086		
Model B	Log(scar count/area)	Std Residuals Log(th) Log(L)	Log(Mass)	AvgTh/GM	Log(BlTh/GM)	Remove Std Res
R^2 values	0.211	+0.075	+0.082	+0.029	+0.024	+0.003

Assessing the Value of the Model

Histograms of the predicted values of both models compared to the histogram of the actual relative removal proportion suggested that the models do not predict reduction particularly well. The 95% confidence interval was calculated for each model using the standard error of the estimate. Given that we are looking at a proportion, a range from 0.0 to 1.0, the confidence intervals indicate that the models are far from optimal.

Model A: 95% confidence interval = estimate +/- 0.414

Model B: 95% confidence interval = estimate +/- 0.415

Assessment of the means of relative removal for each quarter and half of the reduction did not differ significantly from their respective expected means (Table 3). The quarters do significantly differ from each other ($F(3,280) = 1296.935$, $p < 0.001$, Figure 2), as do the first and second halves of reduction ($t(282) = -29.142$, $p < 0.001$, Figure 3).

Table 3: The means of relative removal for each quarter and half do not differ significantly from the expected values.

Group	Mean of relative removal	Expected mean	Degrees of freedom	p-value
Quarter 1	0.129	0.125	69	0.643
Quarter 2	0.367	0.375	80	0.335
Quarter 3	0.623	0.625	63	0.802
Quarter 4	0.874	0.875	68	0.941
Half 1	0.257	0.25	150	0.551
Half 2	0.753	0.75	132	0.800

Figure 2

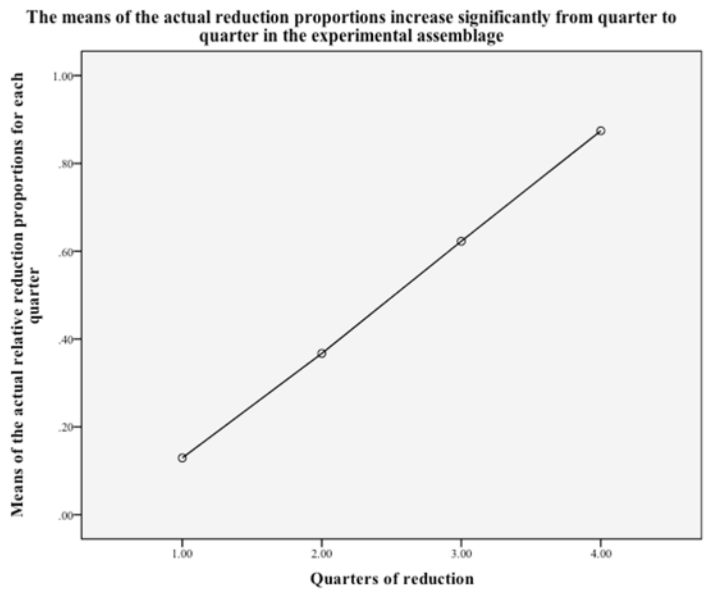
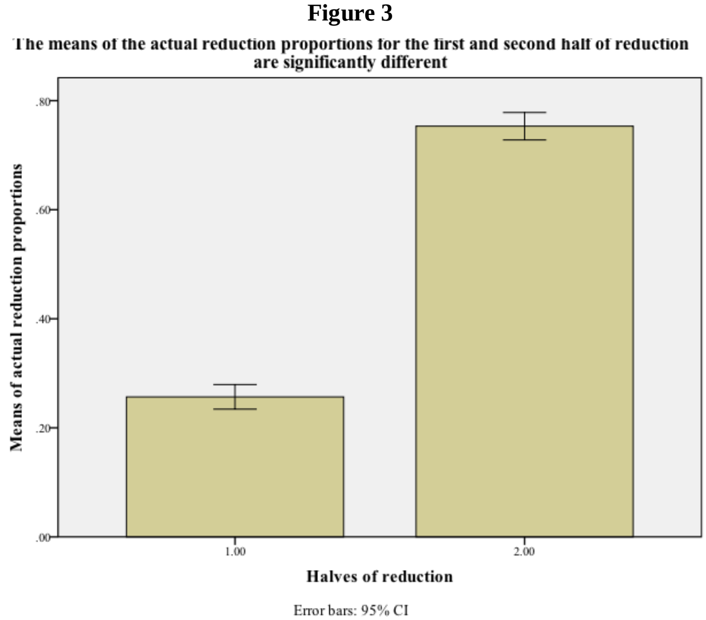


Figure 3



The mean of the predicted values for model A did not differ significantly from 0.50 ($t(171) = -1.113$, $p = 0.267$), nor did they for model B ($t(171) = -1.122$, $p = 0.268$). The means of the predicted values for each quarter of reduction were compared to each other via a one-way ANOVA and post-hoc Tukey HSD tests for both model A and B. For model A, the ANOVA was significant ($F(3,168) = 41.505$, $p < 0.001$), but Tukey HSD tests revealed that while the means for quarter 1 and quarter 2 differ significantly from those of all other quarters, quarter 3 and 4 did not differ from each other significantly ($p = 0.716$, Figure 4). The same was found for model B ($F(3,168) = 41.732$, $p < 0.001$, quarter 3 vs 4 $p = 0.770$, Figure 5). The means of the predicted values for the first half and second half of reduction were compared to each other using an independent samples t-test and found to be significantly different for both models (Model A: $t(141.062) = -10.505$, $p < 0.001$; Model B: $t(140.537) = -10.546$, $p < 0.001$). Comparisons of the means of the predicted values for model A and model B to the means of the actual relative removal proportion for each quarter and half showed that they were significantly different in every case (Table 4).

Figure 4

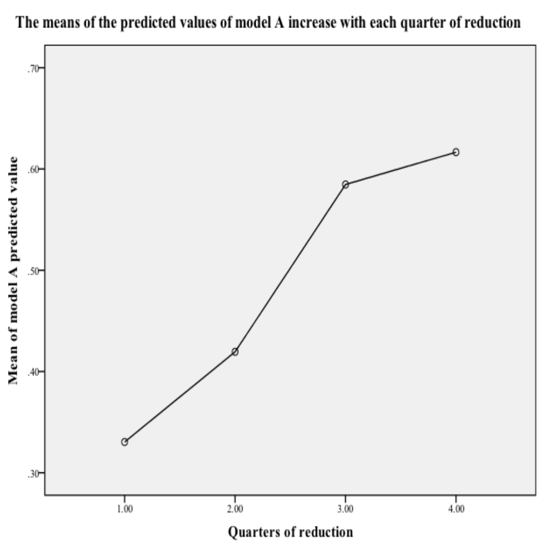


Figure 5

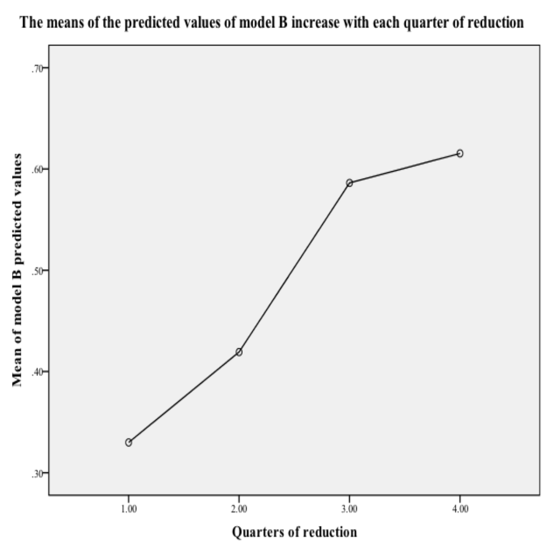


Table 4: Comparisons of the means of the predicted values for each quarter and half to the actual means for each quarter and half show a significant difference.

Group	Actual mean	Model A mean	Model B mean	Degrees of freedom	Mod. A p-value	Mod. B p-value
Quarter 1	0.129	0.33	0.33	44	< 0.001	< 0.001
Quarter 2	0.367	0.419	0.419	41	< 0.001	< 0.05
Quarter 3	0.623	0.585	0.586	47	< 0.05	< 0.05
Quarter 4	0.874	0.617	0.615	36	< 0.001	< 0.001
Half 1	0.257	0.373	0.373	86	< 0.001	< 0.001
Half 2	0.753	0.599	0.599	84	< 0.001	< 0.001

Testing of Assumptions

To ensure that the identified models would be safe to apply, the assumptions of linear regression were tested.

Model A

Normality of residuals was tested using the Shapiro-Wilk test, and it was found that distribution of residuals is normal (Shapiro-Wilk Statistic(172) = 0.991, p = 0.329). Homogeneity of variances was visually assessed from a scatter plot of predicted values versus residuals (Appendix B, Figure B10). The residuals are not homogenous, with variance in residuals increasing as predicted value increases. Linearity of relationships between each predictor variable and relative removal was visually assessed via scatter plots (Appendix A). All relationships were found to be approximately linear. Collinearity statistics (Tolerance and VIF; Table 5) indicate that predictors are not collinear, which is desirable. All "tolerance" values are above 0.10 and all "VIF" values are below ten. When predictors are collinear, coefficient estimates become less reliable. No patterns were found in scatter plots of each of the independent variables versus the residuals (Appendix B, Figures B1 through B4), which means that the assumption of independence is satisfied. In sum, all assumptions except homogeneity of residuals are satisfied for model A.

Table 5: Collinearity statistics show that none of the variables used in model A are collinear.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	2.052	.184		11.171	.000		
	LOGsc_area	.579	.068	.851	8.469	.000	.333	2.999
	logMass	.279	.050	.587	5.589	.000	.305	3.277
	AvgTh_Gm	-1.172	.229	-.331	-5.110	.000	.803	1.245
	LogBlbTh_GM	-.540	.154	-.209	-3.501	.001	.941	1.063

a. Dependent Variable: rel_rem

Model B:

Normality of residuals was tested using the Shapiro-Wilk test, and it was found that distribution of residuals is normal (Shapiro-Wilk Statistic (172) = 0.991, p = 0.360). Like for model A, residuals were not homogenous for model B (Appendix B, Figure B11). All relationships between predictor variables and relative removal were found to be approximately linear (Appendix A). As in model A, no variables are collinear in model B (Table 6). No patterns were found in scatter plots of each of the independent variables versus the residuals (Appendix B, Figures B5 through B9), which means that the assumption of independence is satisfied. Just like for model A, all assumptions except homogeneity of residuals were met for model B.

Table 6: Collinearity statistics for model B

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	T	Sig.	Tolerance	VIF
1	(Constant)	2.013	.205		9.813	.000		
	LOGsc_area	.574	.070	.843	8.221	.000	.322	3.107
	Standardized Residual	-.013	.029	-.042	-.431	.667	.356	2.810
	logMass	.278	.050	.586	5.563	.000	.305	3.279
	AvgTh_Gm	-1.086	.306	-.306	-3.550	.001	.454	2.201
	LogBlbTh_GM	-.502	.178	-.195	-2.814	.005	.708	1.413

a. Dependent Variable: rel_rem

Application

Model A was elected for the application because its prediction value was very close to that of model B, while using fewer variables. However, for the experimental set, three thickness and width measurements were taken for each flake, while the Boxgrove data set only contained one each width and thickness measure. Because of this, the variables used to create model A were recomputed for the experimental data set using only the variables available for the application. The regression was then run again. The R-square value and standard error of the estimate value did not change greatly ($R^2 = 0.410$, std. error of the estimate = 0.213). The newly created predictive equation is as follows:

Model A2:

1.616 + 0.543*log(scar count/area) + 0.253*log(mass) – 0.55*(thickness/geometric mean) – 0.953*log(bulb thickness/geometric mean)

A one-way ANOVA was then run on the quarters of the predicted values from the experimental assemblage for model A2 to ensure that each quarter was still significantly different ($F(3,177) = 40.846$, $p < 0.001$). However, post-hoc Tukey HSD tests revealed that while quarters 1 and 2 differ significantly from every other quarter, quarters 3 and 4 do not differ from each other significantly ($p = 0.885$). An independent samples t-test was performed to compare the mean of the first half of reduction to that of the second, finding that they are significantly different ($t(149.215) = -10.311$, $p < 0.001$). The means for each quarter and half were recorded (Table 7).

Table 7: Reference means created from predicted values of the experimental assemblages

Group	**Mean of predicted value from experimental set**
Quarter 1	0.344
Quarter 2	0.430
Quarter 3	0.592
Quarter 4	0.613
Half 1	0.386
Half 2	0.601

The mean of the predicted reduction proportions for the Boxgrove assemblage was found to be 0.305 (SD = 0.183). This would suggest that the Boxgrove assemblage consists of flakes from the first quarter of reduction. Comparing the histogram of the predicted reduction proportions for the Boxgrove assemblage to that of the predicted values for the experimental set suggests that Boxgrove represents the earlier part of reduction (Figure 6 and Figure 7).

Figure 6

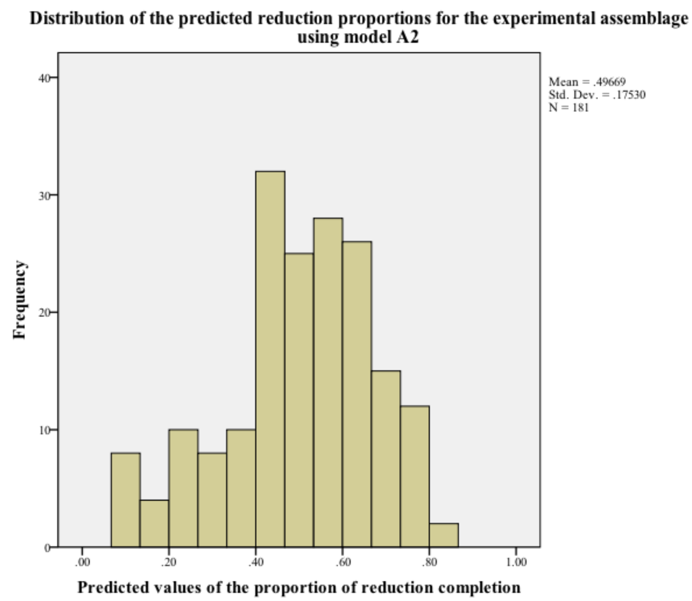
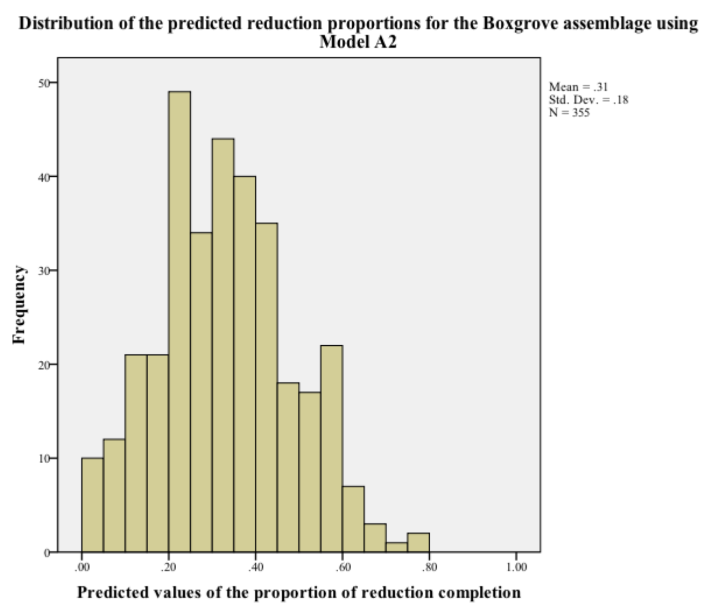


Figure 7



Discussion

The purpose of this study was to create a reduction completion model for refined Acheulean handaxes made from one specific material type, flint. Two potential models were identified, each having a 95% confidence interval of the estimated value +/- 0.414 and 0.415 respectively. The adjusted R^2 values for each were 0.434 and 0.435. When regressed against the actual reduction proportion values, both achieved a slope of 1, which indicates that they do not consistently over or under-predict reduction position.

Model A:

2.052 + 0.579*log(scar count/area) + 0.279*log(mass) – 1.172*(average thickness/geometric mean) – 0.54*log(bulb thickness/geometric mean)

Model B:

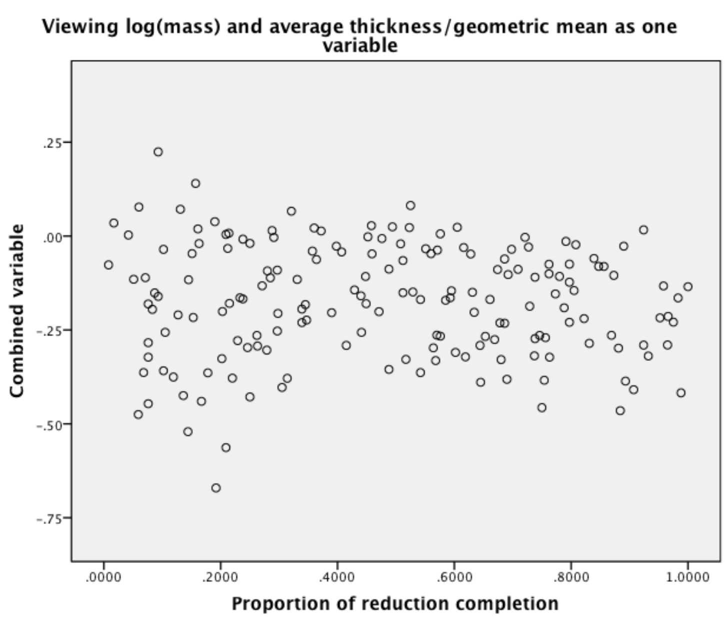
2.013 + 0.574*log(scar count/area) – 0.013*standardized residuals + 0.278*log(mass) – 1.086*(average thickness/geometric mean) – 0.502*log(bulb thickness/geometric mean)

When taking a closer look at each variable used in the regressions, one notices that all but two variables used are controlled for overall size. By dividing by the geometric mean, the bulb thickness and average thickness measurements are controlled for overall flake size. Unfortunately, log(mass), which contributes significantly to the regression, is completely a reflection of flake size. Given how little the standardized residuals of the log(average thickness) versus log(flake length) regression contribute to model B once the other variables have been added, it is not going to be extensively considered here. On its own, the residuals do have a relationship with reduction completion. As reduction completion increases, the standardized residuals decrease in size. The implications of model A will be considered here in more detail.

Scar count divided by area has a positive slope and a high independent R^2 value (0.214). This makes sense, as this is the pattern seen in the individual scatter plot and would imply that scar density increases as reduction progresses. Contrary to what one would expect, the coefficient of the log(mass) variable is positive. One would expect it to be negative given that flake size decreases as the reduction proportion increases. On its own, log(mass) is negatively correlated with reduction progression (Appendix A, Figure A5). This implies that the variable of log(mass) must be counteracting an overestimation caused by another variable.

The last two slopes, those of log(bulb thickness/geometric mean) and average thickness/geometric mean are both negative. This matches the patterns seen in their individual scatter plots (Appendix A, Figures A13 and A16) and means that relative flake thickness and relative bulb thickness decrease as reduction progresses. Based on the scatter plots for log(mass) and average thickness/geometric mean, it appears that the relationship between average thickness/geometric mean and reduction completion is not always well represented as a linear relationship, but that controlling for log(mass) may create a more linear relationship. When these two terms of the model are plotted together as one variable versus reduction completion, the relationship becomes more linear (Figure 8).

Figure 8



In assessing whether the created models meet the assumptions of multiple linear regression, it was found that all assumptions but homogeneity of residuals were met. As the predicted value of reduction completion increased, variance of residuals increased. This suggests that the reliability of the models becomes more unpredictable at higher predicted values, which may make application to archaeological assemblages difficult.

These models should not be viewed as a sole means of analysis. However, when applied to an appropriate archaeological assemblage, they may be useful in, at minimum, differentiating between halves and quarters of reduction, if not providing a more detailed understanding of what parts of reduction occurred at a site. Even with only a broader categorization of first half or second half of reduction, it is possible to begin understanding the function and history of a site. Such a model would be useful because with it we can begin to understand the decisions that contribute to resource management, and how raw-material source location and the geographic region surrounding such an area affect resource economy.

Even if the model had a higher R^2 level and perfectly met all assumptions of multiple linear regression, there would be substantial limitations to the study. Firstly, this study was designed to create a model applicable only to flint assemblages of refined Acheulean handaxe flakes, limiting the reach of its applicability. However, this limitation is necessary on the path to creating effective, material-specific models that supply valuable information. As mentioned in the introduction, the study also faces multiple potential limitations on a more general level.

Firstly, all tools are not created from blanks of the same size. This leads to an inherent size problem, which would require using only computed variables that control for overall size to solve. While three of the four variables used in model A achieve this, the final variable, log(mass), does not. Another way of correcting for this would be to correct for size differences during the application of a model to an assemblage, rather than during the creation of the model itself. However, without knowing the size distribution of the entire assemblage or knowing the size of the original core, correction for size becomes very difficult. Because only parts of a reduction sequence are assumed to be present at a given archaeological site, we could not simply compare the mean sizes for the archaeological versus experimental assemblages, which represent complete reduction, and correct for the difference. Additionally, average differences in size between an archaeological assemblage and the experimental assemblage could change over the course of reduction, which would make correcting for size even more difficult.

The areas of potential limitations discussed in the introduction include lack of repeatable measurements, core material, hammer use, handedness, knapper experience level and personal style, and identification of separate assemblages. Ideally, the issue of repeatable measurements is solved by providing clear descriptions of the measurements taken and the computation of variables. The most inconsistent variable used in the reduction completion model is likely that of scar count. However, by defining a minimum scar length and a clear explanation of the process of identifying scars, discrepancies caused by the scar count measurement are ideally limited. As previously mentioned, the issue of core material is resolved by creating a model that is explicitly to be used for flint assemblages. Skewing caused by knapper handedness is unlikely to significantly affect the variables used in the model. Hammer type was chosen by the knapper throughout the experimental reduction, resulting in a pattern that ideally is relatively equatable to hammer choice during the creation of an actual archaeological assemblage.

The mixing of flakes from differentially experienced knappers is a slightly more worrisome issue. The identified models can only safely be applied to assemblages created by expert knappers. However, past studies have indicated potential methods of differentiating between flakes created by expert versus novice knappers. For example, flake thickness, proper and effective use of platform preparation, and the ability to create relatively longer flakes all change with level of expertise (Stout et al., 2014; Nonaka et al., 2010). In the case of an application, one could first use such methods to differentiate between novice and expert flakes, then apply the percent completion model to the expert flakes. It would be interesting to see whether the portion of reduction identified in the expert flakes of an assemblage is related to the level of novice versus expert flakes found at the site. Perhaps a higher proportion of novice flakes are more likely to be found at sites where the early parts of tool production are being completed because such sites are likely to be closer to raw-material sources, meaning that raw-material is more dispensable for practice by novices.

Finally, the issue of mixed assemblages at large may at times be a problem inhibiting application to archaeological assemblages. This is most likely to be an issue when different reduction types (unifacial and bifacial, for example) are mixed together. However, when all flakes at a site are created during expert bifacial handaxe reduction, mixing of assemblages within those categories is not as large of an issue as it may first seem. While one will not be able to place the flakes within a specific sequence referring to a single reduction, it would still be possible to investigate overall trends at a site.

Comparison to Past Studies

Other reduction studies identified similar variables while creating percent completion models. All three continuum based studies discussed in the introduction used some version of dorsal scar count (Ingbar et al., 1989; Shott, 1996; Bradbury & Carr, 1999). Each of the following variables were used in the reduction completion model of at least one past study: log(thickness), log(area), log(weight), log(platform width), facets, and log(maximum width). In initial linear regressions, all of these variables – except facet number, which was not recorded here – were identified as potential predictor variables. Two of these six variables were directly used in the here identified models, log (thickness) and log (weight). There seems to be general consistency among past studies and this study regarding which flake features are valuable in predicting proportion completion.

Comparing the quality of the past and the newly identified models is quite difficult. While the models here identified had a substantially lower adjusted R^2 value than several of the previously identified models (approximately 0.4 rather than in the 0.8 range like the other studies), this means very little because of differences in methodology. All past continuum studies included a forced intercept of zero, which does not make logical sense and falsely boosts the R^2 value (Casella, 1983). The models prioritized in this study were ones that did not force the regression through the origin. One should not expect the regression to pass through zero because of the high variability in potential first flakes. This variability would make it highly unlikely for the measured attributes of “flake zero” in each reduction to create a predicted value of zero when plugged into the reduction model. Additionally, only whole flakes were included here, which leaves out partial flakes and shatter. The size cut-offs involved in determining a whole flake additionally make it unlikely that a regression involving a size variable (log(mass)) would pass through zero. It may be better to compare the slopes found when the predicted values for each model were regressed with the actual values. The new models achieved a slope far closer to 1.0 when compared with actual reduction values. Additionally, all of the previously identified models relied very heavily on size as an indicator of reduction. None of the variables used in past studies, except when dorsal scar count was divided by flake area or mass, controlled for overall flake size.

In comparison to past studies, this study is limited to one material type and one technology type, giving it the potential to be more powerful within its niche. Within this technology type, multiple experimental assemblages were used, providing a wider sample and more statistical power than several of the other past studies have had. Additionally, unlike in other studies, hammer type was decided fluidly throughout reduction, causing no oddly unnatural patterns of hammer choice. There is evidence that both hard and soft hammers were used at Boxgrove, making it important that both be used in the experimental assemblages (Stout et al., 2014). While varying hammer choice is more naturalistic, it also creates problems because it is very possible that Acheulean knappers did not have the same decision making skills as modern knappers.

In sum, this study substantially adds to the literature by creating a specific and potentially effective model. This study identified the variable of log(bulb thickness/geometric mean) to be significantly related to reduction progression, which has not previously been used or identified. While not able to perfectly predict reduction proportion, this is the first attempt to create a predictive model that does not depend on size. This model enables effective categorization of an assemblage into different portions of an overall reduction sequence. Attempting to use fewer indicators tied to overall size could eventually lead to a model that is far more widely applicable to archaeological assemblages with fewer limitations. Finally, this study has reconfirmed predictor variables identified in past studies.

Boxgrove Application

Boxgrove Background

Boxgrove is located in West Sussex, England, about 12km north of the current shoreline of the English Channel (Roberts & Parfitt, 1999). The site was exposed due to quarrying at ARC Eartham Quarry and excavation began in 1982 and has continued since (Roberts & Parfitt, 1999). The site itself is a middle Pleistocene Acheulean site with high quality preservation of both lithics and fauna (Pope & Roberts, 2005). The site is dated to fall within the range of 524 to 420kyr before present and Oxygen Isotope Stage 13. Correlative mammalian biostratigraphy, specifically using the transition of the water vole, was highly important in the dating of Boxgrove. The site spans an interglacial period and the beginning of the following glacial phase, called the Anglian cold stage (Roberts & Parfitt, 1999).

The site is located near the coast of the English Channel along the base of chalk cliffs. The coastal geography has changed over time, moving through stages of intertidal flats, grasslands, and freshwater marsh (Gamble, 1999). The 20km long chalk cliff runs along the northern edge of the area and is a source of raw flint nodules (Pope & Roberts, 2005). Along this cliff, evidence has been found of test flaking, where knappers began flaking a nodule to assess its quality and abandoned it if undesirable (Pope & Roberts, 2005).

Boxgrove is renowned for its high quality preservation, with several knapping scatters perfectly preserved (Pope & Roberts, 2005). This makes it possible to much more effectively envision where behaviors occurred and understand hominid behavior at the site. While several of the excavated areas represent moments in time, others represent larger spans of time. The main type of lithic tool found at the site is ovate flint handaxes, many of which were found in association with the remains of butchered mammals. These mammal bones had both cut and bite marks on them, with the bite marks overlaying the cut marks, suggesting that hominids had primary access to the carcasses (Pope & Roberts, 2005). A robust tibia that is attributed to *Homo* cf *heidelbergensis* was also found at the site (Gamble, 1999). The tibia is the oldest identified hominid fragment found in the British Isles (Roberts, Stringer, & Parfitt, 1994). Two hominin incisors were also found at the site during 1995 – 1996 (Hillson et al., 2010).

As mentioned, the main type of lithic reduction identified at Boxgrove has been Acheulean bifacial reduction. It was found that the majority of bifaces showed no evidence of retouching or resharpening and that they were discarded shortly after their first usage (Gamble, 1999). The rate of discard decreased as distance from the cliffs increased, which is to be expected. However, the furthest distance from the cliffs that archaeological material was found at was 250m, which is a very short distance and should not alone have caused such a tapering off in handaxe discard (Pope, 2004). Pope and Roberts found that certain sites seemed to be hotspots of hominin activity and contained many discarded handaxes, whereas one-time mammal kill sites rarely had a discarded handaxe associated with it (2005). They suggest that this is because of a difference in ‘mobile’ and ‘fixed’ resources. Handaxes are discarded at sites with fixed resources, such as freshwater, whereas they are discarded infrequently in association with mobile resource sites, such as the site of a mammal kill (Pope & Roberts, 2005). Pope highlights that the availability of freshwater together with proximity to the raw material source would make frequent reuse of a site likely. Area 4c of quarry 1 is an example of this, having once been a seasonal waterhole (Pope & Roberts, 2005).

The assemblage used for this study comes specifically from Quarry 1, Area B, Project D, which is approximately 60m south of the cliff-line (Roberts & Parfitt, 1999). The total area excavated in Quarry 1, Area B is $120m^2$, separated into four trenches. Each trench is 5m by 6m (Roberts & Parfitt, 1999). The flakes identified in Quarry 1/B were categorized as ~20%

finishing flakes, ~12% thinning flakes, ~4% roughing out flakes, and over 60% broken, unidentifiable flakes. Given this information, the area was interpreted as late reduction.

Model A2 Application

Application of model A to the Acheulean Boxgrove assemblage indicates that the flakes in the assemblage fall mainly within the first quarter, or at most the first half, of reduction. The flakes and handaxes of the Boxgrove assemblage are relatively smaller compared to the experimental assemblage. Log(mass) is added in the model, which means that effectively, the predicted reduction proportion value increases as mass increases, even though flake size actually decreases as reduction progresses. This false increase is counteracted in application to a normal flake by the subtraction of other terms. However, when an assemblage is disproportionately small, a lower mean reduction proportion is likely to be predicted. This is magnified when a flake weighs less than 1g because log(mass) becomes negative. When comparing the means of the log(mass) values of the experimental and archaeological assemblages, the value was profoundly lower in the archaeological assemblage (0.563) than in the experimental assemblage (0.988), while means for several of the other variables were greater or approximately equal to those of the experimental assemblage. This suggests that the effect was driven by the difference in weight. An additional limitation to the application is that scar number may have been counted differently for the Boxgrove flakes than during the creation of the model.

Keeping in mind the identified limitations, the flakes from the Boxgrove sample were predicted to be from the first quarter of reduction. In the past, Roberts and Parfitt categorized the highest percentage of flakes from Quarry 1/B at Boxgrove as “finishing flakes” – which would occur at the end of reduction (Roberts and Parfitt, 1999). The majority of handaxes found in the area were not matched to any flakes via refitting, which is why Roberts and Parfitt suggested that many of them were “made outside of the area of excavation” (1999). As mentioned, Quarry 1/B is located approximately 60m from the chalk cliff (Roberts & Parfitt, 1999). Quarry 2, which was characterized as primarily early reduction, is located over 200m from the cliffs. The prediction of early reduction at Quarry 1/B established in this study contradicts past interpretations of the assemblage (Roberts and Parfitt, 1999). Assuming that the application worked properly and given the distance between Quarry 1 and the raw material source, it would make sense that early reduction would be occurring at the site. What specifically would have led to the use of the Quarry 1/B area for early reduction?

Given that there were a relatively high number of handaxes deposited in the area, it is possible that a freshwater source was located nearby, which would have made it an attractive site for continued reuse (Pope & Roberts, 2005). This together with the proximity to the raw material source and the presence of primarily early reduction could suggest that the area around Quarry 1/B was a heavy use area, from which individuals then transported their partially created tool to the site of a mammal kill. Pope and Roberts discuss the benefits of potentially finishing reduction at a kill site (Pope & Roberts, 2005). By transporting the materials to the kill site and knapping there, hominids could limit the amount of time other animals have access to the kill and are able to save more of the carcass for themselves (Pope & Roberts, 2005). In such a case, the cost of carrying a larger piece of stone to the kill site would probably be outweighed by the benefit of protecting the carcass from scavengers. This interpretation would assume that handaxes are primarily made at the site of butcher and are not prepared prior to a hunt. This does seem to be supported by the frequent evidence of knapping at kill sites and the frequent discard of handaxes (Roberts & Parfitt, 1999). Given that Quarry 2 is further from the cliff line (200m) than Quarry 1/B (60m), one would have expected Quarry 2 to be late reduction rather than Quarry 1/B. It could be very interesting to apply the model to Quarry 2 as well to see what part of reduction the model classifies it as. Further application of reduction proportion models to Boxgrove assemblages and more research on the accuracy of the application would be necessary before further suggestions or conclusions. Currently, the primary concern with the application of the model to Boxgrove is that the effect seems to be largely driven by the difference in size between the archaeological and experimental assemblages.

The existence of a large difference in the mean log(mass) between the Boxgrove assemblage and the experimental assemblages when there is no large difference in the other variables used in the model may suggest that the experimental assemblages were not made in the same way as the archaeological assemblages. This would mean that a method was used by Paleolithic knappers that modern knappers are unfamiliar with. Investigating what the main cause for the difference in size was could lead to interesting findings regarding the techniques used in creating stone tools.

Future Directions

Ideally, this study could be expanded by using additional experimental assemblages. With a larger data set, it is likely that the accuracy of the current model could be increased. Creation of several additional experimental assemblages would also make it possible to test the predictive model on an experimental assemblage that was not used in the creation of the model. This would be one of the most effective tests of the accuracy and applicability of the established model.

As has been previously discussed, the greatest limitation that the study currently faces in application is the use of log(mass). This leads to problems when applying the model to differently sized assemblages. Future research should focus on variables that are controlled for size. If this is not possible, the next step would be to look into ways to correct for the size difference before applying the model to an archaeological assemblage.

Continued application of the model to archaeological assemblages, especially assemblages that are close in size to the experimental assemblages, would be valuable both in terms of information gathered about the site and in terms of testing the applicability of the model. Finally, given that this model could only be applied to refined Acheulean handaxe assemblages of flint, the next step in achieving the big picture goal of investigating resource transportation at many archaeological sites would be to create models for more reduction types and materials.

Conclusion

The goal of this study was to identify an effective, applicable reduction proportion equation for refined Acheulean handaxes made of flint and to apply it to one of the archaeological assemblages from the Boxgrove site in southern England. This was achieved, with scar count/area, average thickness/geometric mean, bulb thickness/geometric mean, and mass being identified as the key contributing variables. This model is a significant contribution to the literature because of its specificity to one type of material and reduction process, and its relatively good ability to categorize an assemblage within the quarters or halves of reduction. When applied to the Quarry 1/B assemblage from Boxgrove, the site was characterized as early reduction, contrary to past interpretations. However, the mean difference in size between the archaeological and experimental assemblages is what drove the effect, suggesting that the interpretation may not be accurate.

Given the surprise of the Boxgrove finding, further testing and research on both the model and its application to Boxgrove and other sites will be crucial. Ideally, the sample size used to create the models could be increased through further experimental replications and the model could be tested upon a further experimental assemblage. The overall effects of using a size dependent variable within the reduction equation also need to be investigated further. On the broadest scale, such reduction models could ideally be created for a variety of materials and technologies, enabling application to a wide range of archaeological sites.

Appendix A

Figure A1: As reduction progresses, interior angle does not change significantly (R^2=0.003, p>0.05).

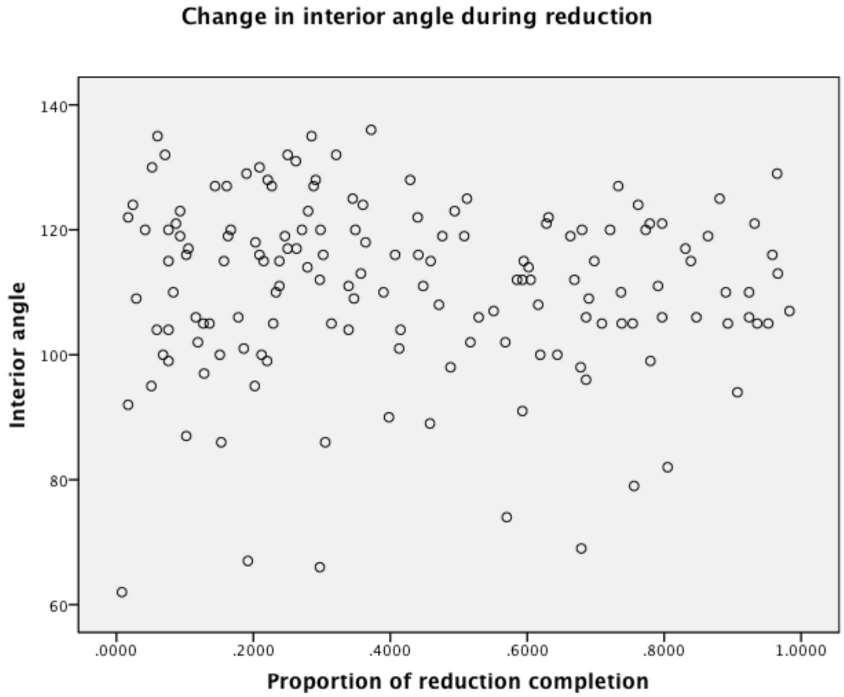


Figure A2: As reduction progresses, exterior angle does not change significantly (R^2=0.005, p>0.05).

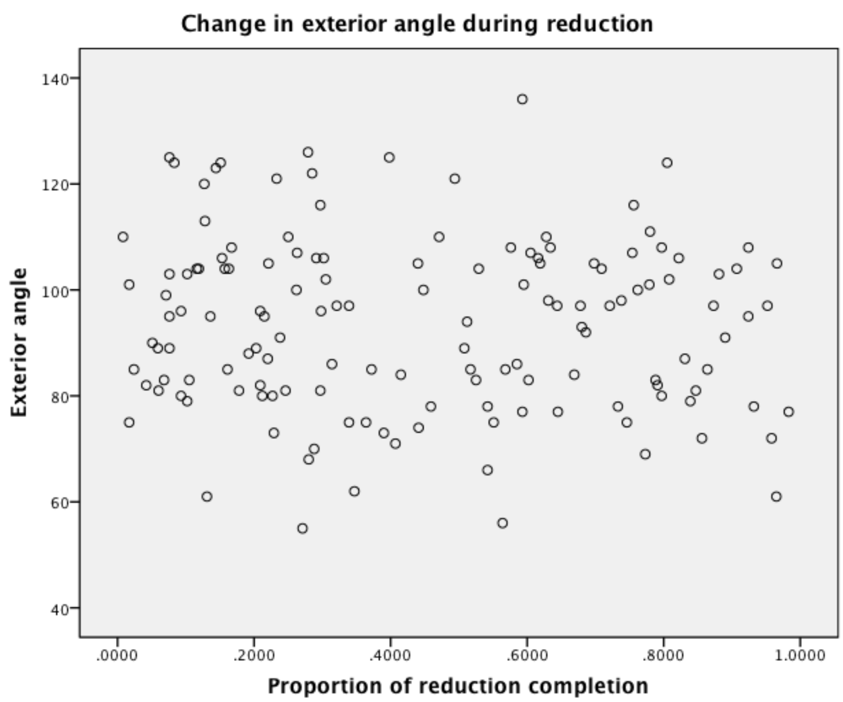


Figure A3: Width decreases as reduction progresses (R^2=0.071, p<0.001).

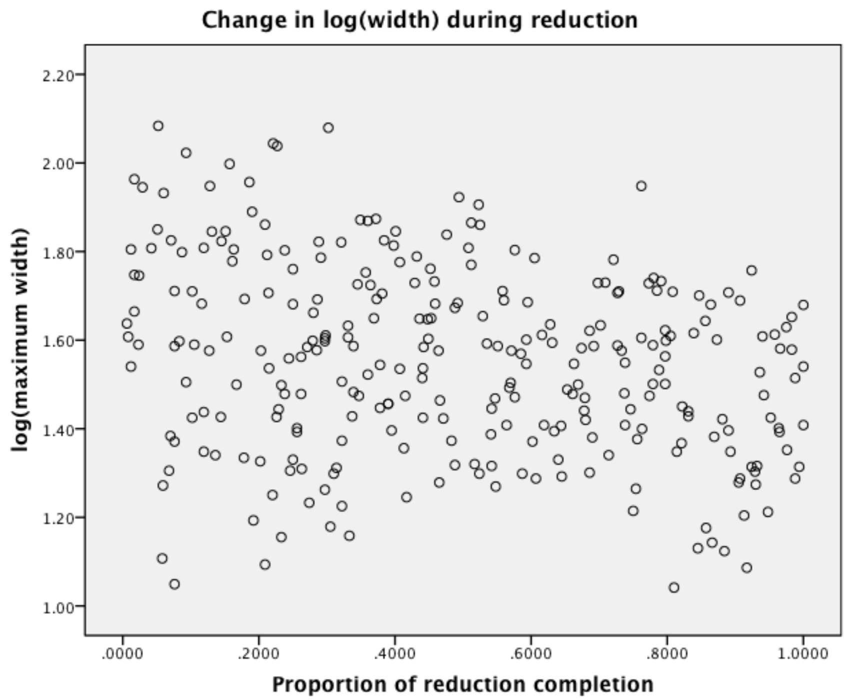


Figure A4: Flake length decreases as reduction progresses (R^2=0.044, p<0.001).

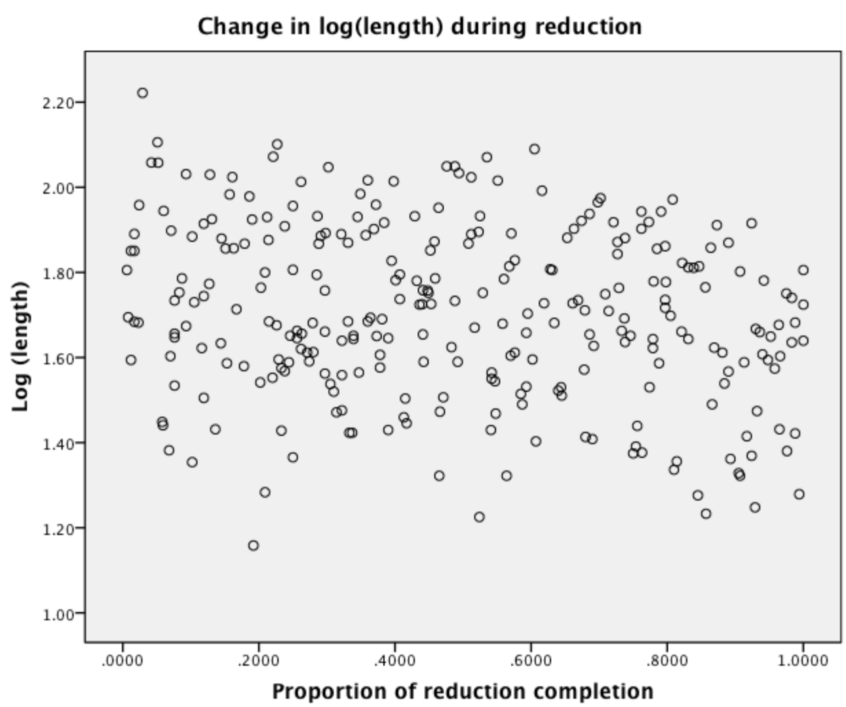


Figure A5: As reduction progresses, mass decreases (R^2=0.102, p<0.001).

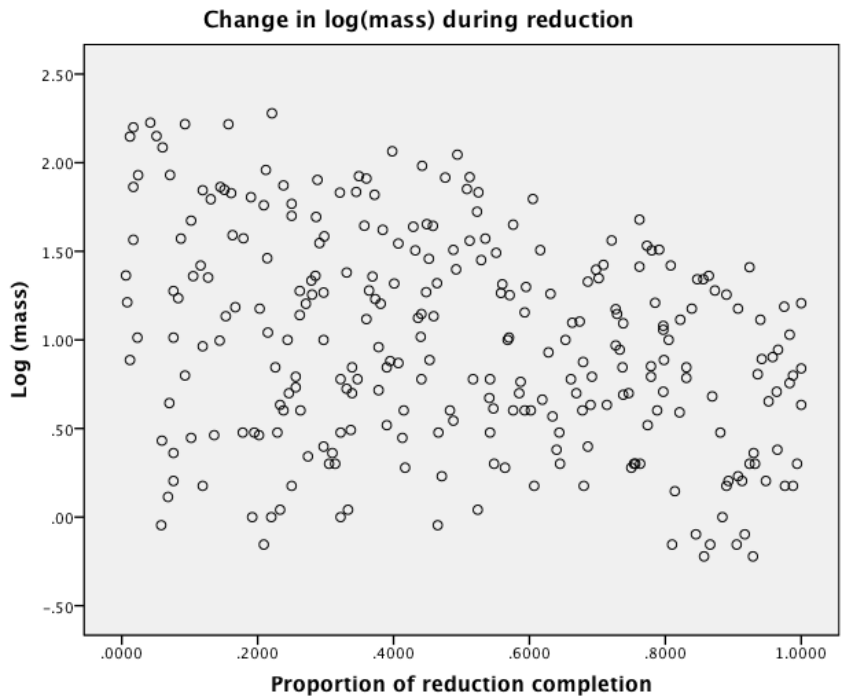


Figure A6: Scar density increases as reduction progresses (R^2=0.214, p<0.001).

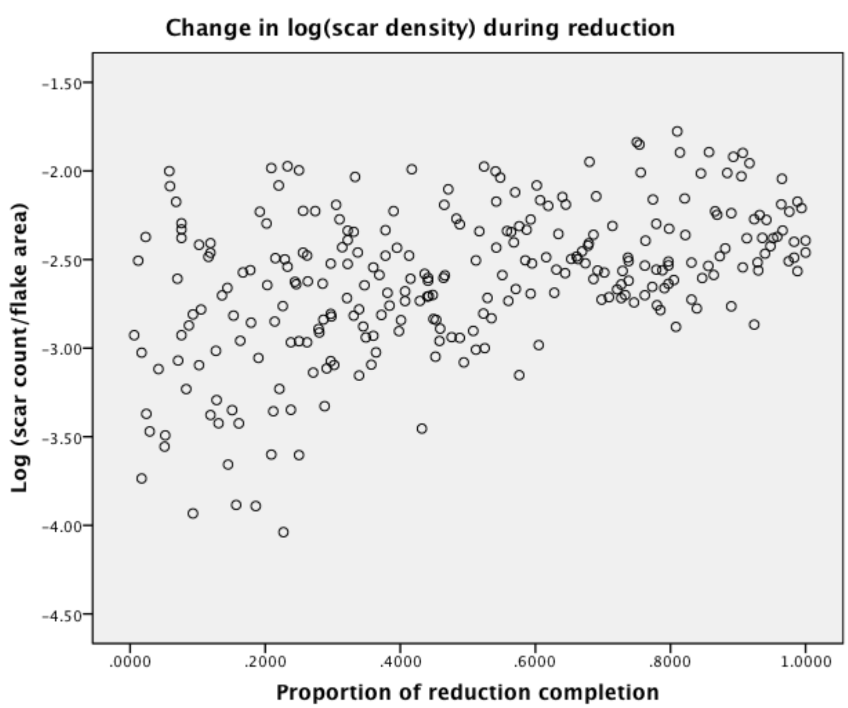


Figure A7: AvgDifTh is here defined as the average of the absolute values of each of the three thickness measurements subtracted from the average thickness for each flake. Flake thickness consistency appears to decrease during reduction (R^2=0.116, p<0.001).

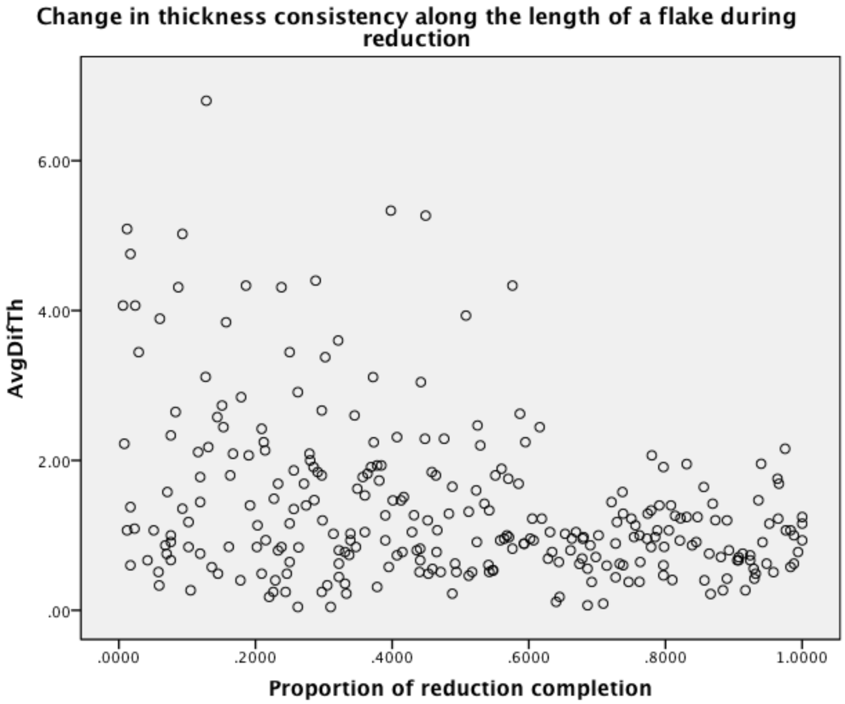


Figure A8: AvgDifW is here defined as the average of the absolute values of each of the three width measurements subtracted from the average width for each flake. Flake width consistency appears to decrease during reduction (R^2=0.041, p<0.001)

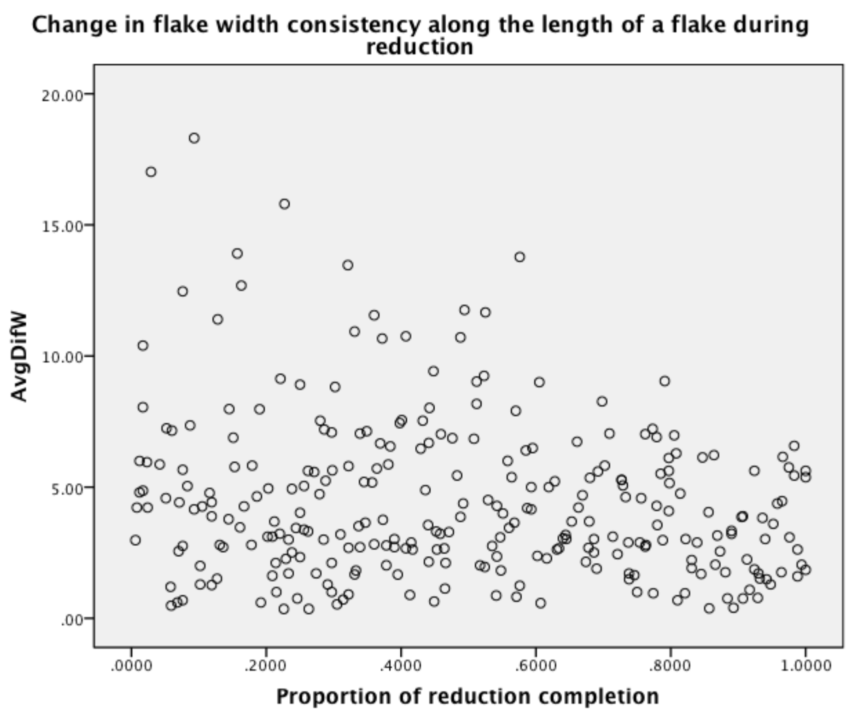


Figure A9: The same definitions were used as in Figure A7. Again, we see flake thickness consistency decrease during reduction (R^2=0.059, p<0.001)

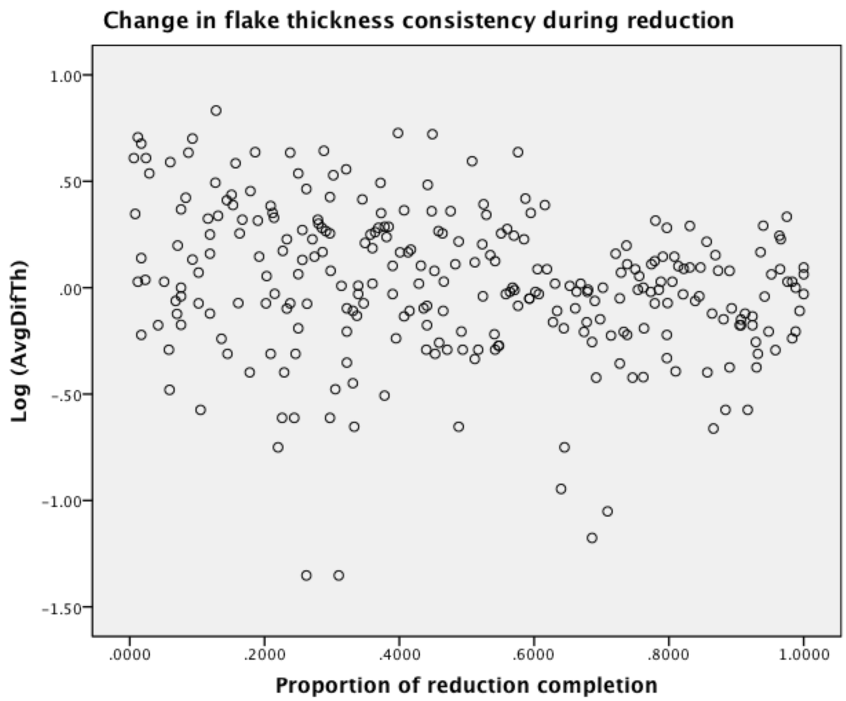


Figure A10: The same definitions were used as in Figure A8. Again, we see flake width consistency decrease during reduction (R^2=0.019, p<0.05).

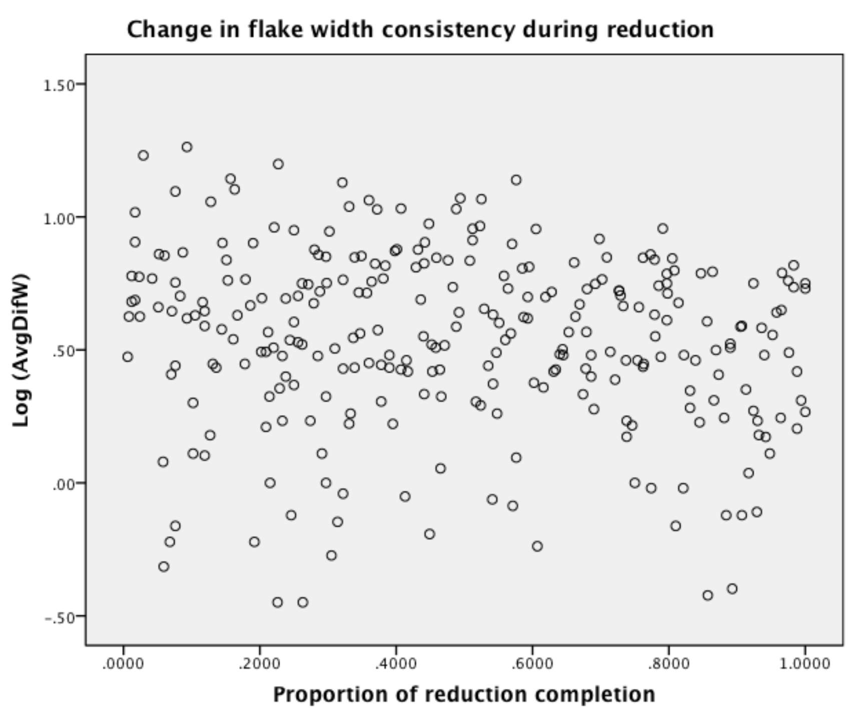


Figure A11: Overall flake size decreases during reduction (R^2=0.082, P<0.001).

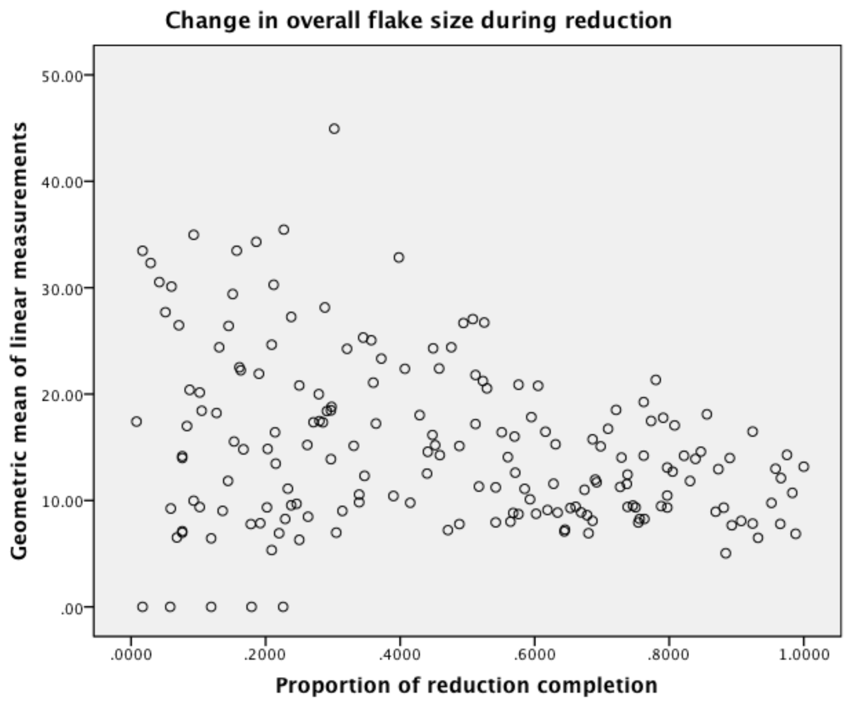


Figure A12: Platform area decreases during reduction (R^2=0.101, p<0.001).

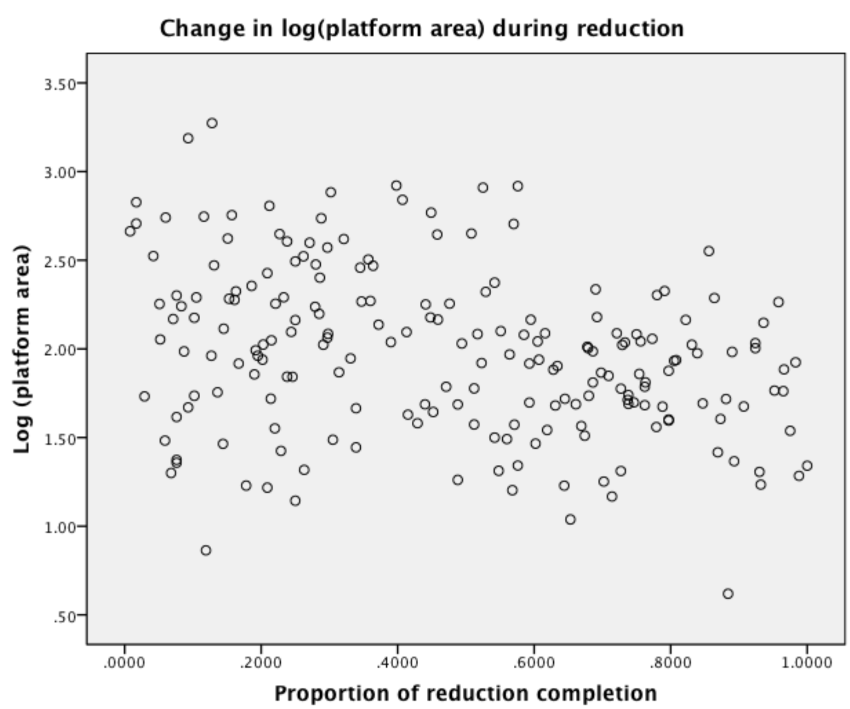


Figure A13: Average thickness controlled for flake size decreases during reduction (R^2=0.144, p<0.001).

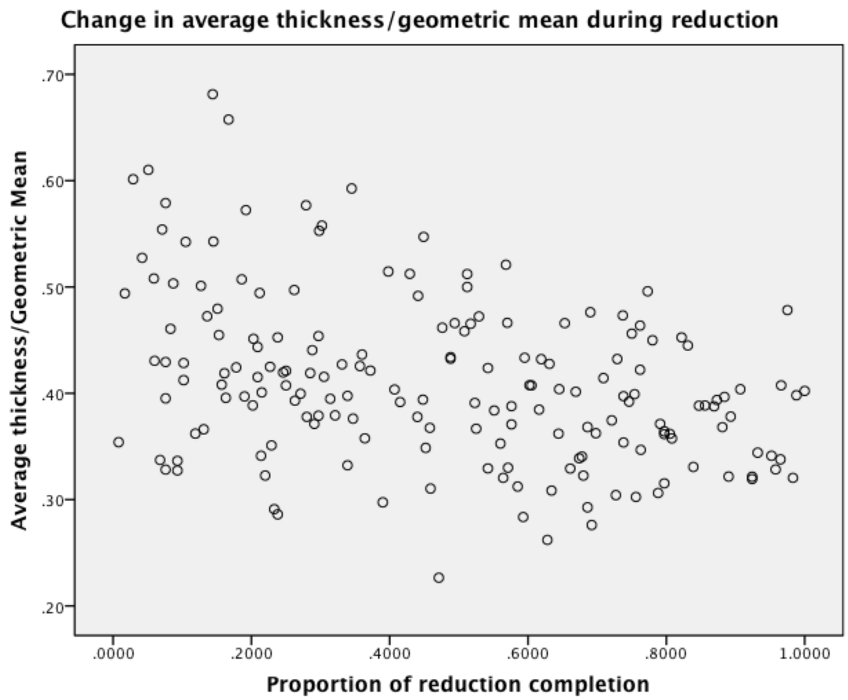


Figure A14: Average width controlled for flake size does not appear to change substantially during reduction, but does increase slightly (R^2=0.078, p<0.001).

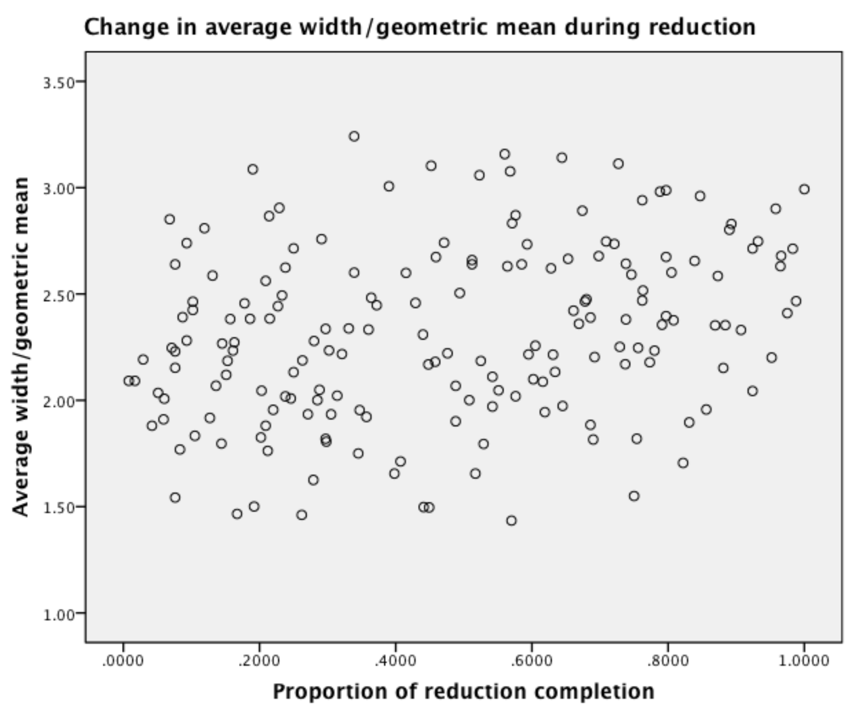


Figure A15: Flake length controlled for overall flake size increases as reduction progresses (R^2=0.10, p<0.001).

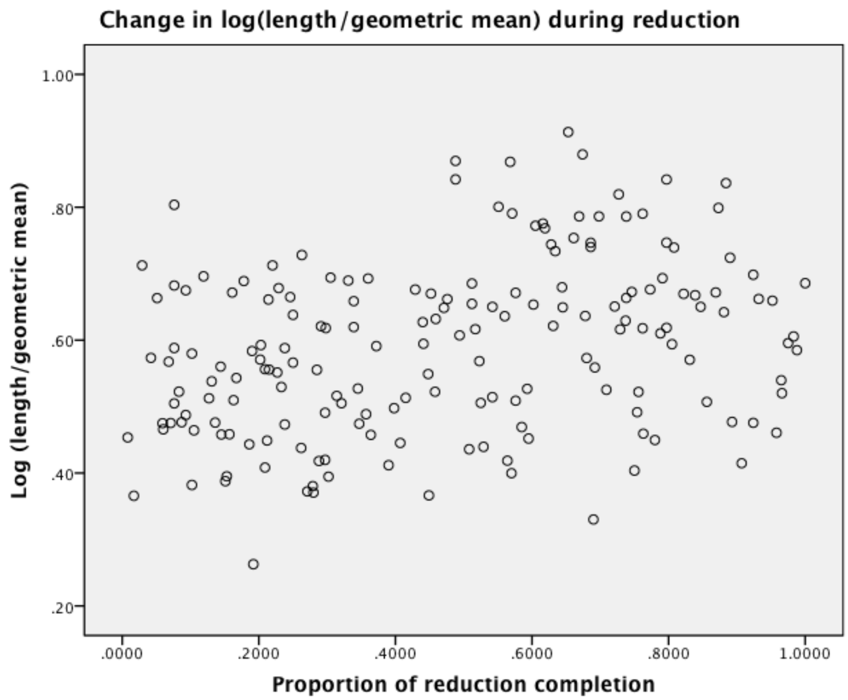


Figure A16: Bulb thickness controlled for overall flake size decreases as reduction progresses (R^2=0.096, p<0.001).

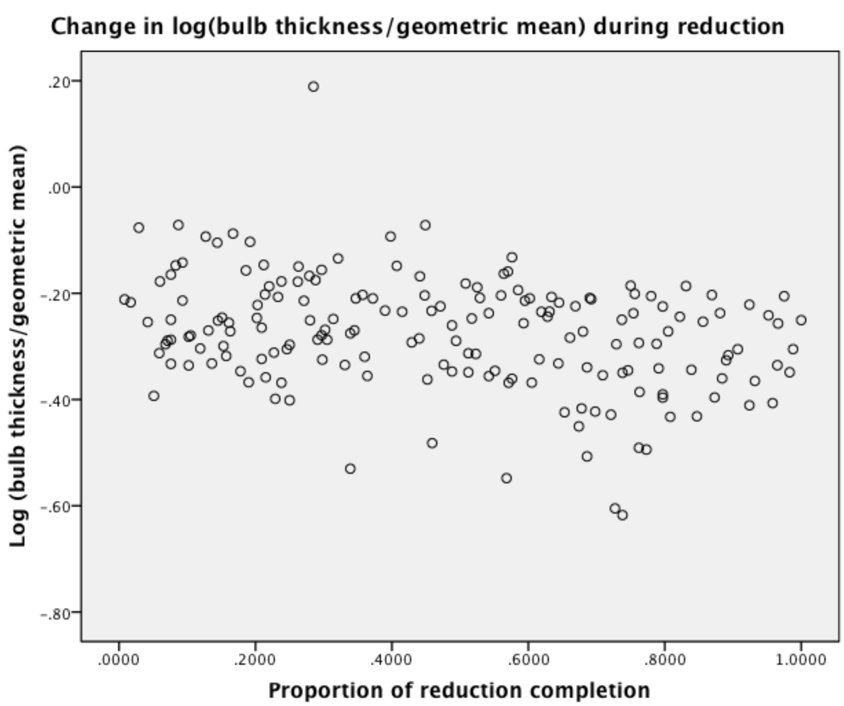
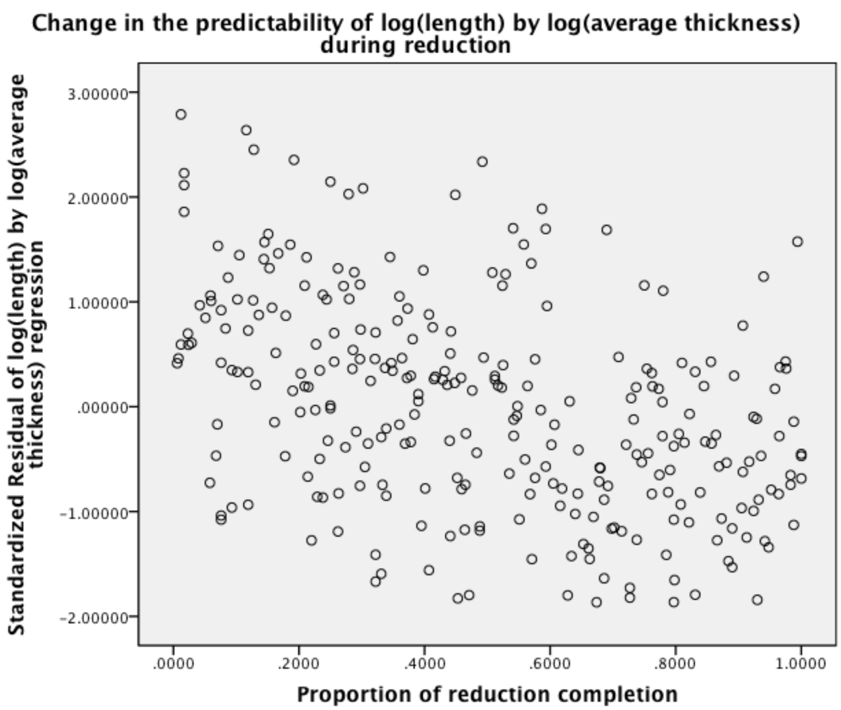


Figure A17: As reduction progresses, a regression predicting log(length) from Log(average thickness) moves from a tendency to positively over-predict length to under-predicting length (R^2=0.193, p< 0.001).



Appendix B

Figure B1

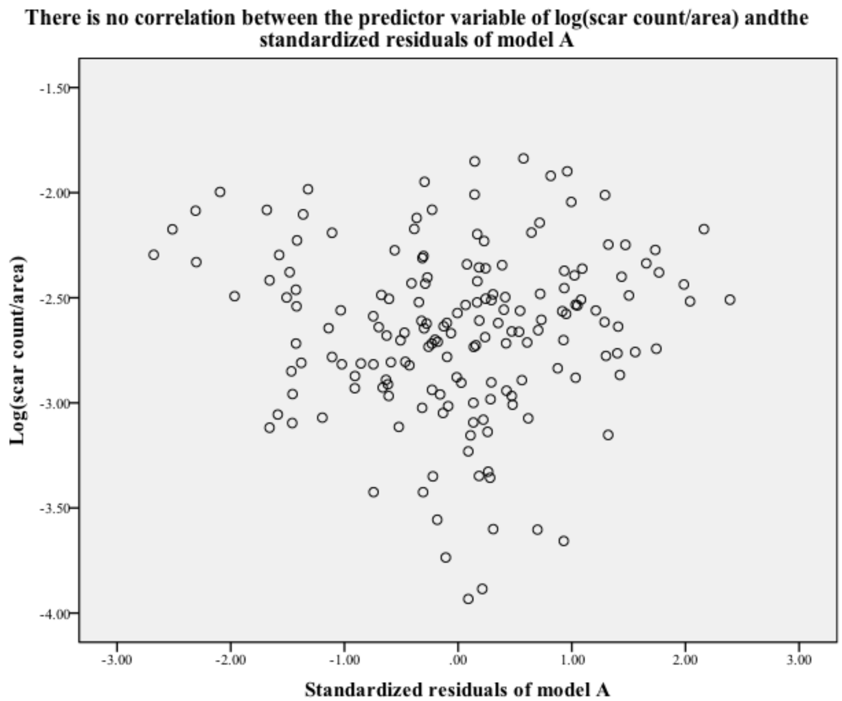


Figure B2

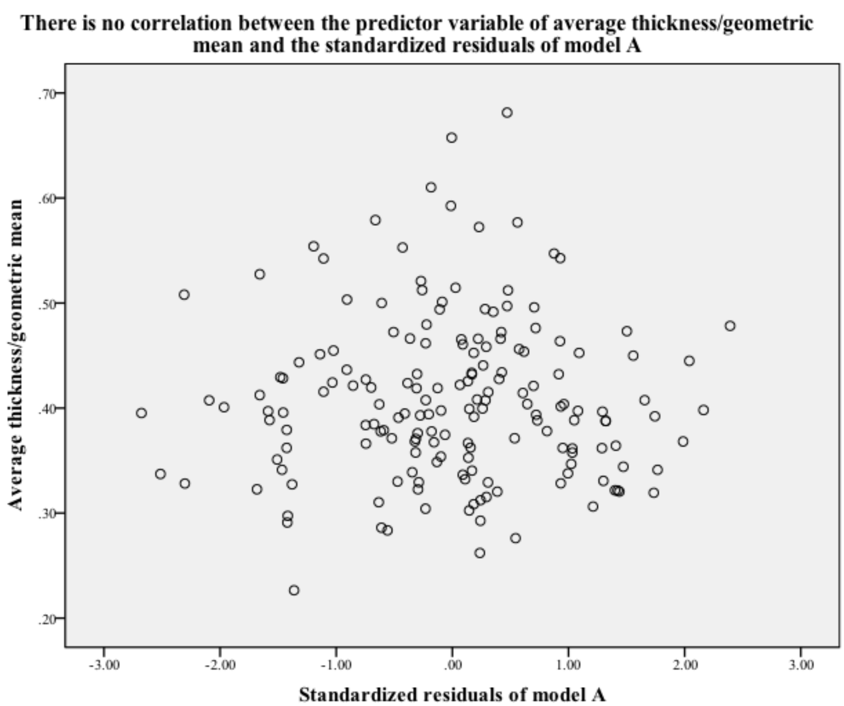


Figure B3

There is no correlation between the predictor variable of log(mass) and the standardized residuals of model A

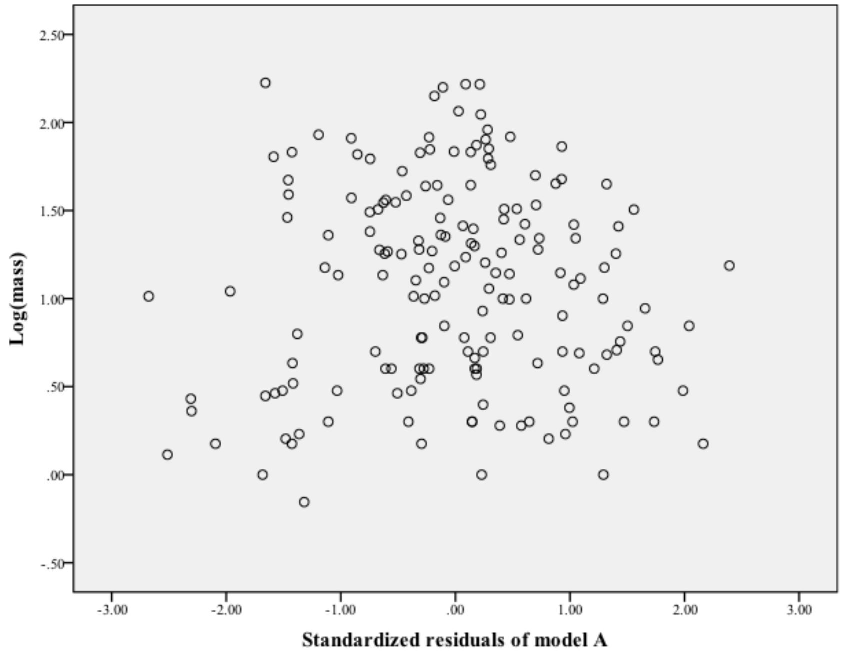


Figure B4

There is no correlation between the predictor variable of log(bulb thickness/geometric mean) and the standardized residuals of model A

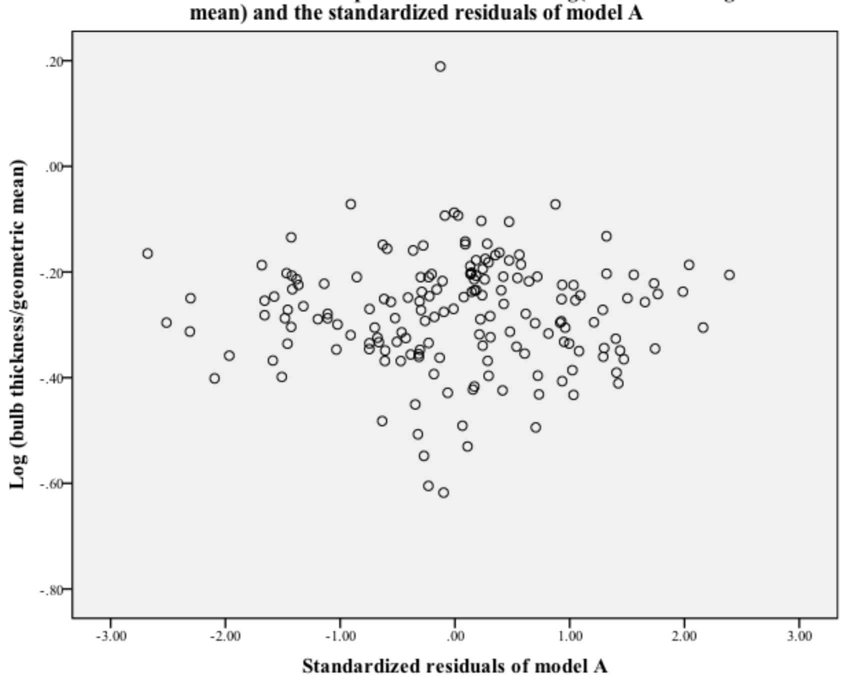


Figure B5

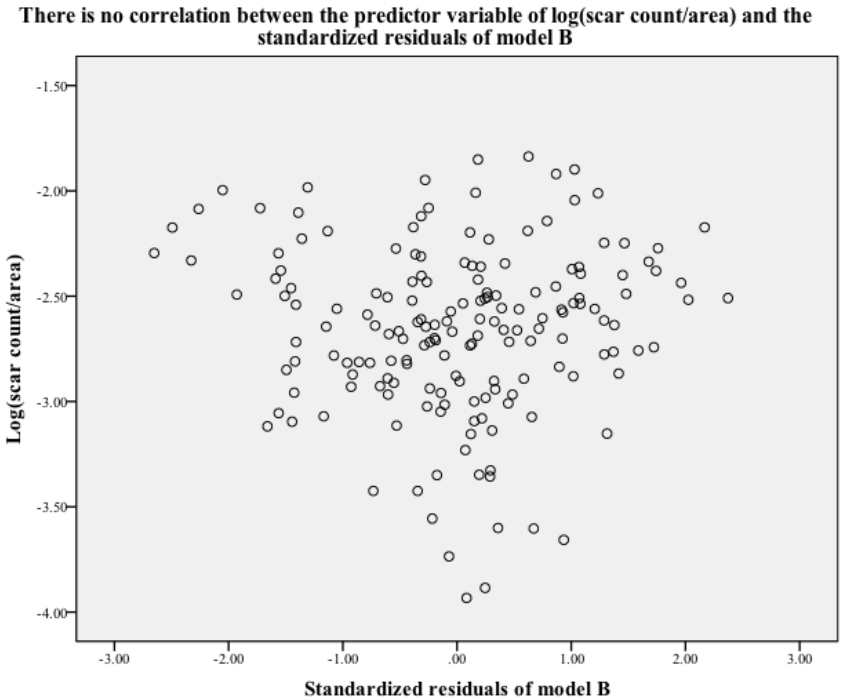


Figure B6

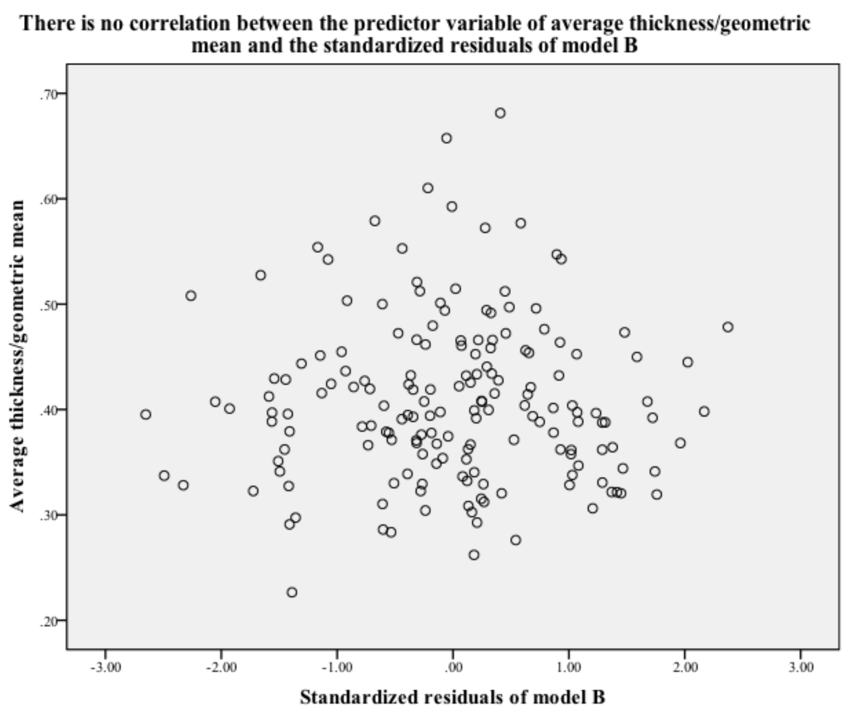


Figure B7

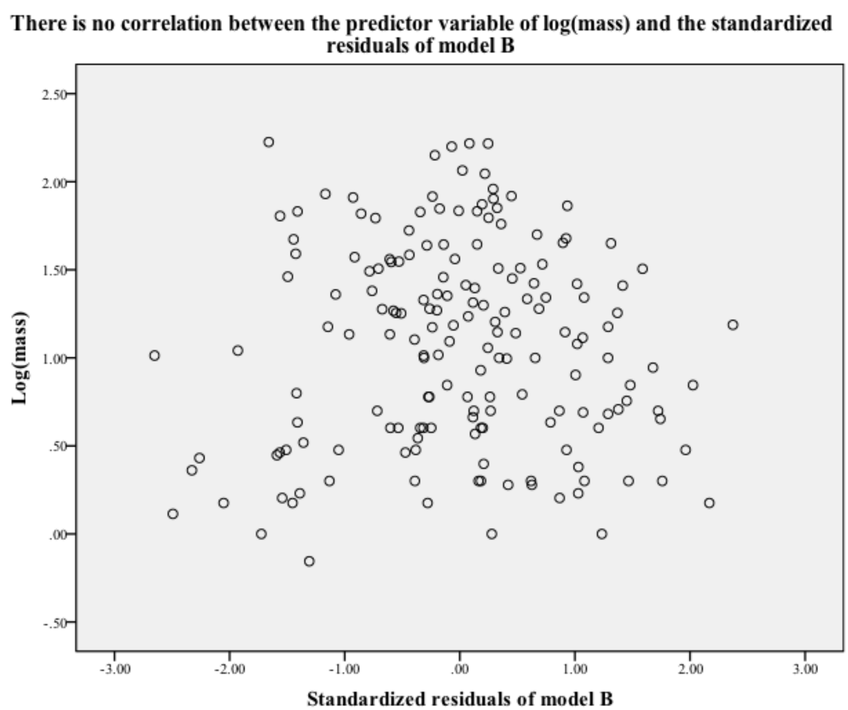


Figure B8

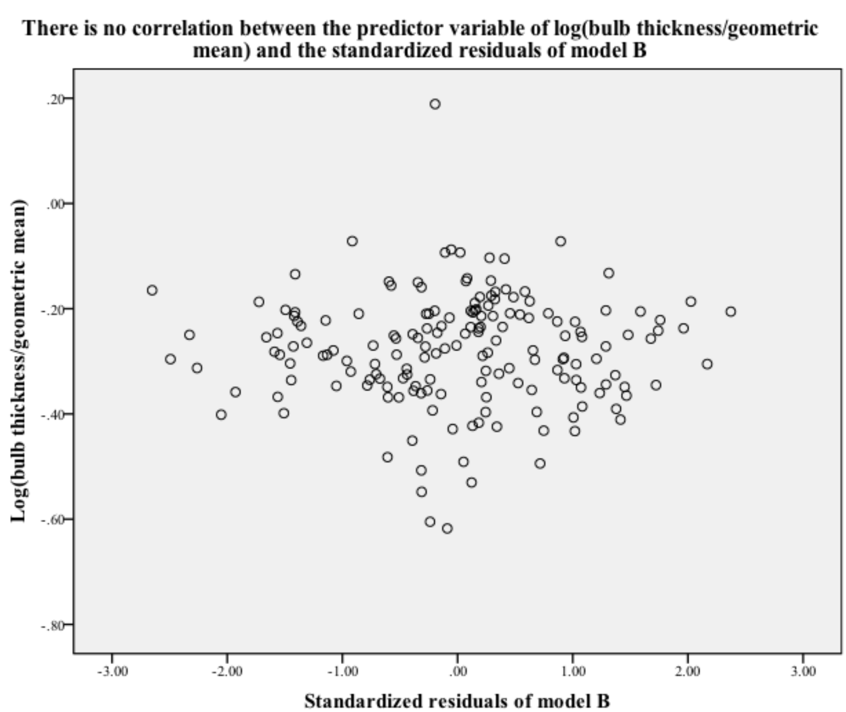


Figure B9

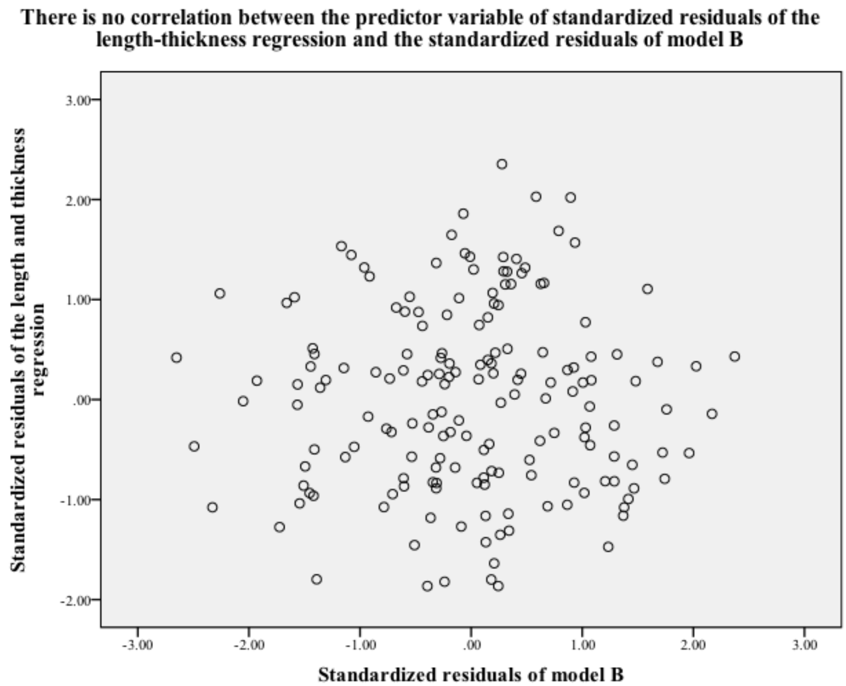


Figure B10

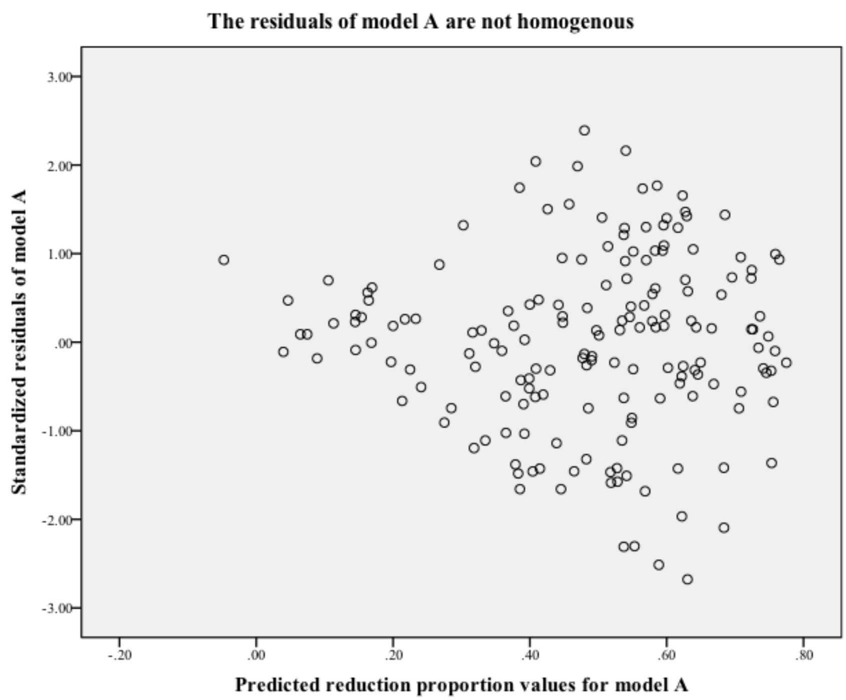
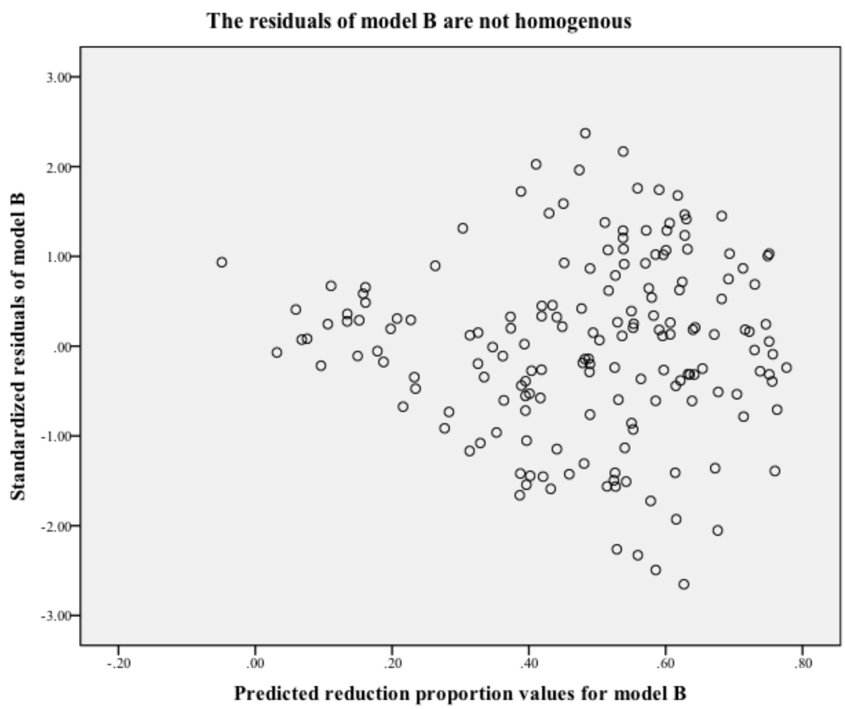


Figure B11


The residuals of model B are not homogenous
3.00
2.00
1.00
.00
-1.00
-2.00
-3.00
Standardized residuals of model B
-.20
.00
.20
.40
.60
.80
Predicted reduction proportion values for model B

Works Cited

Ambrose, S. (2001). Paleolithic Technology and Human Evolution. *Science, 291*(5509), 1748-1753.

Amick, D., & Mauldin, R. (1997). Effects of Raw Material on Flake Breakage Patterns. *Lithic Technology, 22*(1), 18-32.

Amick, D., Mauldin, R., & Tomkin, S. (1988). An Evaluation Of Debitage Produced By Experimental Bifacial Core Reduction of a Georgetown Chert Nodule. *Lithic Technology, 17*(1), 26-36.

Bargalló, A., & Mosquera, M. (2014). Can hand laterality be identified through lithic technology? *Laterality: Asymmetries of Body, Brain and Cognition, 19*(1), 37-63.

Bradbury, A., & Carr, P. (1995). Flake Typologies and Alternative Approaches: An Experimental Assessment. *Lithic Technology, 20*(2), 100-115.

Bradbury, A., & Carr, P. (1999). Examining Stage and Continuum Models of Flake Debris Analysis: An Experimental Approach. *Journal of Archaeological Science,* 105-116.

Braun, D. R., Rogers, M. J., Harris, J. W., & Walker, S. J. (2008). Landscape-scale variation in hominin tool use: evidence from the developed Oldowan. *Journal of Human Evolution, 55*(6), 1053-1063.

Casella, G. (1983). Leverage and regression through the origin. *The American Statistician, 37*(2), 147-152.

Collins, M. (1975). Lithic Technology as a Means of Processual Inference. In *Lithic Technology: Making and using stone tools*.

Cotterell, B., & Kamminga, J. (1987). The Formation of Flakes. *American Antiquity, 52*(4), 675-675.

Debenath, A., & Dibble, H. (1994). Fundamental Concepts and Terms. In *Handbook of paleolithic typology*. Philadelphia: University Museum, University of Pennsylvania.

Ensor, H., & Roemer, E. (1989). Comments on Sullivan and Rozen's "Debitage Analysis and Archaeological Interpretation" *American Antiquity, 54*(1), 175-178.

Gamble, C. (1999). The first European societies 500,000-300,000 years ago. In *The Palaeolithic Societies of Europe* (pp. 119-121). Cambridge, U.K.: Cambridge University Press.

Goodman, M. (1944). The Physical Properties of Stone Tool Materials. *American Antiquity, 9*(4), 415-415.

Hillson, S. W., Parfitt, S. A., Bello, S. M., Roberts, M. B., & Stringer, C. B. (2010). Two hominin incisor teeth from the middle Pleistocene site of Boxgrove, Sussex, England. *Journal of human evolution*, *59*(5), 493-503.

Hiscock, P., & Tabrett, A. (2010). Generalization, inference and the quantification of lithic reduction. *World Archaeology, 42*(4), 545-561.

Hodgson, D., & Mc Nabb, J. (2005). More On Acheulean Tools. *Current Anthropology*, *46*(4), 647-650.

Hopkinson, T., Nowell, A., & White, M. (2013). Life histories, metapopulation ecology, and innovation in the Acheulian. *PaleoAnthropology*, *2013*, 61-76.

Ingbar, E., Larson, M., & Bradley, B. (1989). A Nontypological Approach to Debitage Analysis. In *Experiments in Lithic Technology*. Oxford, England: B.A.R.

Jones, P. (1979). Effects of Raw Materials on Biface Manufacture. *Science, 204*(4395), 835-836.

Kooyman, B. (2000). *Understanding stone tools and archaeological sites*. Calgary: University of Calgary Press.

Laughlin, J. P., & Kelly, R. L. (2010). Experimental analysis of the practical limits of lithic refitting. *Journal of Archaeological Science*, *37*(2), 427-433.

Lycett, S., & Gowlett, J. (2008). On questions surrounding the Acheulean 'tradition'. *World Archaeology, 40*(3), 295-315.

Magnani, M., Rezek, Z., Lin, S. C., Chan, A., & Dibble, H. L. (2014). Flake variation in relation to the application of force. *Journal of Archaeological Science, 46,* 37-49.

Magne, M., & Pokotylo, D. (1981). A pilot study in bifacial reduction sequences. *Lithic Technology, 10*(2/3), 34-47.

Morrow, T. (1997). A Chip Off the Old Block: Alternative approaches to debitage analysis. *Lithic Technology, 21*(1), 51-69.

Nonaka, T., Bril, B., & Rein, R. (2010). How do stone knappers predict and control the outcome of flaking? Implications for understanding early stone tool technology. *Journal of Human Evolution, 59,* 155-167.

Patterson, L. (1990). Characteristics of Bifacial-Reduction Flake-Size Distribution. *American Antiquity*, *55*(3), 550-558.

Pelcin, A. (1997). The Effect of Indentor Type on Flake Attributes: Evidence from a Controlled Experiment. *Journal of Archaeological Science, 24,* 613-621.

Pope, M., & Roberts, M. (2005). Observations on the relationship between Palaeolithic individuals and artefact scatters at the Middle Pleistocene site of Boxgrove, UK. In C. Gamble & M. Porr (Eds.), *The Hominid Individual in Context*. New York: Routledge Taylor and Francis Group.

Proffitt, T., & de la Torre, I. (2014). The effect of raw material on inter-analyst variation and analyst accuracy for lithic analysis: a case study from Olduvai Gorge. *Journal of Archaeological Science*, *45*, 270-283.

Roberts, M. B., & Parfitt, S. A. (1999). *Boxgrove: A Middle Pleistocene Hominid Site at Eartham Quarry, Boxgrove, West Sussex*. English Heritage.

Roberts, M. B., Stringer, C. B., & Parfitt, S. A. (1994). A hominid tibia from Middle Pleistocene sediments at Boxgrove, UK. *Nature*, *369*(6478), 311-313.

Roe, D. (2002). An Overview, With Some Thoughts on the Study of Bifaces. In *Multiple approaches to the study of bifacial technologies*.

Rozen, K., & Sullivan, A. (1989). The Nature of Lithic Reduction and Lithic Analysis: Stage Typologies Revisited. *American Antiquity, 54*(1), 179-184.

Rugg, G., & Mullane, M. (2001). Inferring handedness from lithic evidence. *Laterality: Asymmetries of Body, Brain and Cognition*, *6*(3), 247-259.

Shipton, C., & Clarkson, C. (2015). Flake scar density and handaxe reduction intensity. *Journal of Archaeological Science: Reports*, *2*, 169-175.

Shott, M. (1994). Size and form in the analysis of flake debris: Review and recent approaches. *Journal of Archaeological Method and Theory, 1*(1), 69-110.

Shott, M. (1996). Stage versus continuum in the debris assemblage from production of a fluted biface. *Lithic Technology, 20*(1), 6-22.

Stemp, W. J., Childs, B. E., Vionnet, S., & Brown, C. A. (2009). Quantification and discrimination of lithic use-wear: surface profile measurements and length-scale fractal analysis. *Archaeometry*, *51*(3), 366-382.

Stout, D. (2002). Skill And Cognition In Stone Tool Production: An Ethnographic Case Study From Irian Jaya. *Current Anthropology, 43*(5), 693-722.

Stout, D., Apel, J., Commander, J., & Roberts, M. (2014). Late Acheulean technology and cognition at Boxgrove, UK. *Journal of Archaeological Science*,*41*, 576-590.

Stout, D., & Chaminade, T. (2012). Stone tools, language and the brain in human evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1585), 75-87.

Tostevin, G. (2011). Levels of Theory and Social Practice in the Reduction Sequence and Chaîne Opératoire Methods of Lithic Analysis. *PaleoAnthropology,* 351-375.

Toth, N. (1982). The Stone Technologies of Early Hominids at Koobi Fora, Kenya: An Experimental Approach. Ph.D. Dissertation, University of California, Berkeley.

Toth, N., & Schick, K. (2007). 21 Overview of Paleolithic Archeology. In *Handbook of Paleoanthropology* (pp. 1943-1963).

Uomini, N. (2009). The prehistory of handedness: Archaeological data and comparative ethology. *Journal of Human Evolution*, *57*, 411-419.

Williams, J. P., & Andrefsky, W. (2011). Debitage variability among multiple flint knappers. *Journal of Archaeological Science*, *38*(4), 865-872.

Winton, V. (2005). An investigation of knapping-skill development in the manufacture of Palaeolithic handaxes. *Stone Knapping: The Necessary Conditions for a Uniquely Hominin Behaviour. Cambridge: McDonald Institute for Archaeological Research*, 109-116.

Yerkes, R., & Kardulias, P. (1993). Recent developments in the analysis of lithic artifacts. *Journal of Archaeological Research*, *1*(2), 89-119.