The role of experimental knapping in empirically testing key themes in the evolution of lithic technology: reduction intensity, efficiency and behavioural complexity

Antoine Muller
BA (archaeology)
BA Honours (archaeology)

A thesis submitted for the degree of Master of Philosophy at The University of Queensland in 2017

School of Social Science
Abstract

Experimental knapping has complimented and stimulated lithic analyses for over a century. Throughout this period, the discipline has witnessed an increase in the scientific rigour and theoretical grounding with which these studies are conducted. This thesis charts these key trends and in doing so establishes a best-practice model of experimental knapping, the veracity of which is in turn tested using four new lithic experiments. These case-studies employ experimental knapping to advance our understanding of flake platform measurement, reduction intensity, technological efficiency, and behavioural complexity.

The first case-study, Chapter 3, offers a more accurate and precise calliper-based method of flake platform measurement that relies on simple geometric approximations of platform shape rather than the inflexible and unreliable existing method of multiplying platform width by thickness. In Chapter 4, a new reduction intensity metric for backed blades, a hitherto overlooked tool-type, is developed and tested on the backed blades from an early Neolithic site in Turkey. This new metric allows a reconstruction of the raw material consumption patterns at the site, finding that the backed blades likely contributed to conserving the inhabitants’ scarce lithic raw material. Meanwhile, Chapter 5 outlines the results of a comparison of the raw material efficiency of eight different lithic technologies, finding that lithic technological efficiency was a generally ascending trend over the last 3.3 million years and that the main transition in efficiency occurred between the Lower to Middle Palaeolithic. On a similar time-scale, Chapter 6 explores the behavioural complexity of five different technologies by charting the relative levels of hierarchical organisation required for their production, finding equivalencies in the behavioural complexity required for the tool-kits of Neanderthals and their contemporary Homo sapiens.

These four case-studies, coupled with a consideration of existing knapping experiments, allow an understanding of how experimental knapping is embedded in the broader archaeological research process, and ultimately tests the efficacy of a best-practice model of experimental knapping. This model identifies the initial scope, methodological control, and breadth of interpretations as the key variables dictating the validity of an experiment. While knapping experiments may differ markedly in their scope and control, they do not necessarily vary in their validity. Instead, it is the interplay of these variables that dictates the validity of experimentation. Within this best-practice model, lithic experiments are most robust when the scale of the initial scope, methodological control and ensuing interpretations are congruent, and when they involve explicit and falsifiable hypothesis testing.
Declaration by Author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.
Publications During Candidature

Peer-Reviewed Journal Articles (Published)


Publications Included in This Thesis


<table>
<thead>
<tr>
<th>Contributor</th>
<th>Statement of contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antoine Muller (Candidate)</td>
<td>Experimental design (90%)</td>
</tr>
<tr>
<td></td>
<td>Analysis (90%)</td>
</tr>
<tr>
<td></td>
<td>Writing (95%)</td>
</tr>
<tr>
<td></td>
<td>Editing (80%)</td>
</tr>
<tr>
<td>Chris Clarkson</td>
<td>Experimental design (10%)</td>
</tr>
<tr>
<td></td>
<td>Experimental sample (100%)</td>
</tr>
<tr>
<td></td>
<td>Analysis (10%)</td>
</tr>
<tr>
<td></td>
<td>Writing (5%)</td>
</tr>
<tr>
<td></td>
<td>Editing (20%)</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Contributor</th>
<th>Statement of contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antoine Muller (Candidate)</td>
<td>Experimental design (30%)</td>
</tr>
<tr>
<td></td>
<td>Experimental sample (30%)</td>
</tr>
<tr>
<td></td>
<td>Analysis (90%)</td>
</tr>
<tr>
<td></td>
<td>Writing (90%)</td>
</tr>
<tr>
<td></td>
<td>Editing (80%)</td>
</tr>
<tr>
<td>Chris Clarkson</td>
<td>Experimental design (70%)</td>
</tr>
<tr>
<td></td>
<td>Experimental sample (70%)</td>
</tr>
<tr>
<td></td>
<td>Analysis (10%)</td>
</tr>
<tr>
<td></td>
<td>Writing (10%)</td>
</tr>
<tr>
<td></td>
<td>Editing (20%)</td>
</tr>
</tbody>
</table>

For the list of author contributions relating to the submitted but unpublished articles, refer to the tables at the beginning of Chapters 4 and 6.
Contributions by Others to the Thesis
Zoe Heighway performed a pilot study as part of an Honours project that developed and tested the method used in Chapter 5, and provided a portion of the data used in that chapter. Chapter 5 built on that pilot study by expanding the sample and analyses conducted. Dr Tyler Faith provided assistance with statistical analyses and insightful comments on the manuscripts presented here. Tierney Lu, Kasih Norman, Perri Braithwaite, Kathy Lai, and Jackie Child, participated in the inter-observer variability portion of Chapter 3.

Statement of Parts of the Thesis Submitted to Qualify for the Award of another Degree
Chapter 6 was based on a pilot study conducted by the author (AM), submitted as partial fulfilment of the requirements of an Honours degree at the University of Queensland, 2014. This chapter significantly expanded the sample, methodology, theory, and analytical scope of the pilot study.
Acknowledgements

I would like to sincerely thank my supervisors, Chris Clarkson and Andrew Fairbairn for their unending guidance, feedback and support throughout this process. A special thank you goes to Chris Clarkson for imparting and contributing his abundant expertise in lithic technology and knapping, on which this thesis so often relied. I also greatly appreciate the insightful comments made by the two anonymous reviewers.

This project would not have been possible without the support of a great number of friends and colleagues. Thank you to Tierney Lu, Kasih Norman, Perri Braithwaite, Kathy Lai and Jackie Child for participating in an inter-observer variability study. Thank you also to Maddy Moyle and Kasih Norman who at Boncuklu assisted with artefact analysis and lithic illustration respectively. I am particularly grateful to Zoe Heighway whose pilot study formed the basis for one of the papers contained in this thesis and who kindly provided a portion of the data. A special thank you goes to Jacques Pelegrin for participating in knapping experiments and providing valuable insights and conversations. Thank you also to Ofer Bar-Yosef for valuable conversations throughout this project. I am also deeply grateful to Tyler Faith for patient advice on statistical analyses and for providing insightful comments on the manuscripts in this thesis. Emilija Nicolosi and Chat Marasinghe deserve many thanks as the laboratory technicians at the School of Social Science of the University of Queensland for equipment and logistical support as well as their tireless efforts in maintaining the smooth operation of the labs.

I am also sincerely grateful to the funding bodies that facilitated this research, including an Australian Government Research Training Program Scholarship (RTP), a UQ School of Social Science fieldwork bursary, a UQCHU-UniQuest bursary, and an Australian Archaeological Association Student Research grant. Additionally, excavation at Boncuklu was funded by The British Institute in Ankara, British Academy (Research Development Award BR100077), Australian Research Council (DP120100969), National Geographic award (GEFNE 1-11) and University of Oxford (Wainwright Fund). Thank you also to the directors of the Boncuklu excavations, Douglas Baird, Andrew Fairbairn and Gökhan Mustafaoğlu, who allowed access to the lithic assemblage.

Finally, to my friends and family, who offered an unending and unconditional supply of support and encouragement, I owe a debt of gratitude. Without you, none of this would be possible. This thesis is dedicated to the memory of Keith Morton.
Keywords
experimental archaeology, lithic technology, knapping, human evolution, technological efficiency, reduction intensity, Palaeolithic, Neolithic, stone tools

Australian and New Zealand Standard Research Classifications (ANZSRC)
ANZSRC code: 210102 Archaeological Science, 80%
ANZSRC code: 210103 Archaeology of Asia, Africa and the Americas, 10%
ANZSRC code: 210105 Archaeology of Europe, the Mediterranean and the Levant, 10%

Fields of Research (FoR) Classification
FoR code: 2101 Archaeology, 100%
Table of Contents

Abstract .......................................................................................................................................... ii  
Declaration by Author ..................................................................................................................... iii  
Publications During Candidature .................................................................................................... iv  
Publications Included in This Thesis ............................................................................................... v  
Contributions by Others to the Thesis ............................................................................................ vi  
Statement of Parts of the Thesis Submitted to Qualify for the Award of another Degree .......... vi  
Acknowledgements ........................................................................................................................ vii  
Keywords ...................................................................................................................................... viii  
Australian and New Zealand Standard Research Classifications (ANZSRC) ......................... viii  
Fields of Research (FoR) Classification ........................................................................................ viii  
Table of Contents ........................................................................................................................... ix  
List of Figures ................................................................................................................................ xii  
List of Tables ................................................................................................................................ xiii  
List of Abbreviations ...................................................................................................................... xiv  
CHAPTER 1: Introduction ............................................................................................................... 1  
  1.1 Introduction............................................................................................................................ 1  
  1.2 Research Questions .............................................................................................................. 3  
  1.3 Thesis organisation ............................................................................................................... 3  
CHAPTER 2: Literature Review and Background............................................................................ 5  
  2.1 Introduction............................................................................................................................ 5  
  2.1.1 Experimentation as Middle-Range Theory Construction.................................................. 6  
  2.1.2 Experimentation and the Scientific Method ..................................................................... 8  
  2.2 History of Knapping Studies ................................................................................................ 10  
  2.2.1 Early Research ............................................................................................................. 10  
  2.2.2 Criticisms of These Early Studies .................................................................................. 11  
  2.2.3 Addressing Criticisms: Hypothesis Testing and Recent Trends ..................................... 13  
  2.2.4 Recent Applications of Experimental Knapping ............................................................. 14  
  2.3 Towards a Best-Practice Approach to Experimental Knapping ............................................ 19  
  2.3.1 A Model of Experimental Knapping ............................................................................... 19  
  2.3.2 Designing a Knapping Experiment: Methodological Control .......................................... 21  
  2.3.3 Designing a Knapping Experiment: Sample Size .......................................................... 25  
  2.3.4 Designing a Knapping Experiment: Relevance to the Archaeological Record ............... 26  
  2.4 Case Studies and Hypothesis Testing ................................................................................. 26  
  2.4.1 Hypothesis Development: Chapter 3 ............................................................................. 27  
  2.4.2 Hypothesis Development: Chapter 4 ............................................................................. 27  
  2.4.3 Hypothesis Development: Chapter 5 ............................................................................. 28  
  2.4.4 Hypothesis Development: Chapter 6 ............................................................................. 28  
  2.5 Summary ............................................................................................................................. 29
List of Figures

**Figure 2.1:** Model representing the traits of theoretically valid knapping experiments 21

**Figure 3.1.** Photographs of platforms superimposed with the four geometric approximations 35

**Figure 3.2.** Boxplot comparing the rectangular approximations with the actual closest approximations and the predicted closest approximations 38

**Figure 3.3.** Boxplot comparing the rectangular approximations with all other approximations 40

**Figure 3.4.** Boxplot comparing the rectangular approximations with the halved rectangular approximations, actual closest approximations, and the predicted closest approximations 41

**Figure 3.5.** Regressions of individual measurements conducted by six observers plotted against the mean of those measurements 44

**Figure 4.1.** Schematic drawing of a blade’s plan and side view demonstrating where the three width, thickness and edge angle measurements were recorded 54

**Figure 4.2.** Scatter plots showing the strongest regressions of the relationship of blade blank thickness to length and width 57

**Figure 4.3.** Scatter plot of actual original blank mass versus predicted original blank mass 58

**Figure 4.4.** Drawings of chipped stone from Boncuklu 61

**Figure 4.5.** Histograms showing the reduction intensity of the experimental and Boncuklu backed blades 62

**Figure 4.6.** Bivariate boxplot (‘bagplot’) showing the morphology of Boncuklu blade blanks and microliths 64

**Figure 4.7.** Stacked bar charts showing the presence and location of platform preparation 67

**Figure 5.1.** Photograph of a blank and two stages in the process of reducing the photograph to a measurable polygon 82

**Figure 5.2.** Bar chart showing the cutting edge per gram values for each technology 84

**Figure 5.3.** Bar chart showing the cutting edge per gram values for each time period 85

**Figure 5.4.** Scatter plots examining influences on the cutting edge length per gram of core for individual flakes 87

**Figure 5.5.** Boxplots of cutting edge per gram of original core for different platform, termination, and platform preparation types 89

**Figure 6.1.** Photographs of the replicative assemblage 96

**Figure 6.2.** Definitions of each component of the hierarchical diagrams 103

**Figure 6.3.** Hierarchical diagram of the cognitive processes of bipolar knapping 106

**Figure 6.4.** Hierarchical diagram of the cognitive processes of discoidal knapping 106
Figure 6.5. Hierarchical diagrams of the cognitive processes of handaxe knapping

Figure 6.6. Hierarchical diagrams of the cognitive processes of Levallois knapping

Figure 6.7. Hierarchical diagrams of the cognitive processes of blade core knapping

Figure 6.8. Clustered bar chart showing the hierarchical breadth and depth of each diagram

Figure 6.9. Box-plots showing the frequency of phase changes for each reduction sequence

Figure 6.10. Bar chart showing the mean hierarchical breadth throughout each reduction sequence

Figure 7.1. A synthesis of the range of sources of evidence that contribute to hypothesis generation in knapping experiments

Figure 7.2. The best-practice model for experimental knapping developed in Chapter 2 with each case-study overlain

List of Tables

Table 3.1. Descriptions of each of the different geometric approximations used in this study

Table 3.2. Summary data for all different measurement sets

Table 3.3. Measures of spread among the different measurement types

Table 3.4. Results of both measures of inter-observer variability

Table 4.1. Summary data for the experimental and Boncuklu specimens

Table 4.2. Regressions of the relationship of blade thickness to length and width

Table 4.3. Counts of blanks, tools and cores analysed thus far at Boncuklu.

Table 5.1. Mass values of the bipolar, multiplatform, discoidal, biface, Levallois, prismatic blade, punch blade and pressure blade cores, waste and blanks from each reduction sequence

Table 5.2. Mean mass, counts and cutting edge values for each reduction strategy

Table 5.3. Number of blanks, mass of waste and cutting edge per gram values between the expert knapper and intermediate knapper for each technology

Table 6.1. Summary of the 21 reduction sequences conducted, including figure numbers and quantitative features of the hierarchical diagrams

Table 7.1. Summary of the archaeological relevance of each case-study
List of Abbreviations

2D – Two Dimensional

3D – Three Dimensional

BK – Boncuklu

Cal. B.C.E. – Calibrated (date) Before Common Era

CV – Coefficient of Variation

df – Degrees of Freedom

EPA – External Platform Angle

GAP – Geometric Approximations of Platforms

IQR – Inter-Quartile Range

kya – Thousand Years Ago

LSA – Late Stone Age

mya – Million Years Ago

N – Number (sample size)

SD – Standard Deviation

TEM - Technical Error of Measurement
CHAPTER 1:  
Introduction

1.1 Introduction  
A fundamental problem faced by archaeologists is reconstructing past human behaviour from a fragmentary, static and taphonomically biased archaeological record. Analogy, the use of presently observable phenomena as a proxy for the unobservable, allows archaeologists to reconstruct these dynamic behaviours from the material remains of the past. One of the most powerful means of analogy creation is experimentation. While experiments have long served archaeology in general, this thesis specifically explores the role of experimentation in our understanding of lithic technology. With analogy forming the foundation of much archaeological research, and with experimentation one of only a few available avenues of analogy creation in archaeology, the reliability of experimentation is fundamental to the reliability of archaeology itself. This thesis therefore explores means of ensuring the validity of lithic experiments. The primary aim of this thesis is to develop a best-practice model of experimental knapping and test the efficacy of this model using four case-studies comprised of new lithic experiments. These experiments develop methods of lithic measurement (Chapter 3) and test hypotheses surrounding reduction intensity (Chapter 4), technological efficiency (Chapter 5), and behavioural complexity (Chapter 6).

Lithic experimentation has a long history and has become an increasingly common and powerfully applied approach to archaeology. Accordingly, experimental knapping is the subject of several syntheses that attempt to summarise its history and theoretical background (Bradley 1977; Carr and Bradbury 2010; Eren et al. 2016; Flenniken 1984; Johnson 1978; Lamdin-Whymark 2009; Lerner 2010; Nami 2010; Olausson 2010). The four case studies contained in this thesis are therefore not designed to encompass the entire range of lithic experimentation. Rather, they show a portion of the various approaches and possible applications. These experiments have variable degrees of scope and methodological control. Most criticisms of knapping experiments attempt to diminish their validity based on concerns with these initial parameters (Hayashi 1968; Shea 2011; Thomas 1986). For example, detractors highlight that some experiments are established with such a wide breadth of scope that ensuing interpretations serve as mere generalisations accompanied by countless exceptions. Others argue that a portion of knapping experiments are too loosely controlled to be meaningful, and that some are too tightly controlled to have bearing on the archaeological record. These criticisms foreground the importance of establishing appropriate experimental parameters. While this sentiment is shared here, it is argued that a single parameter alone is insufficient to
evaluate the validity of an experiment. Instead, it is posited that the interplay between the scope and methodological control influences the scale of interpretations. Knapping experiments are open to criticism where interpretations violate the scale of these initial parameters. Accordingly, a model is put forth in the following chapter that synthesises the relationships among these variables. Additionally, the importance of the hypothetico-deductive model of scientific reasoning is outlined in the following chapter and it is argued that explicit and falsifiable hypothesis testing is a prerequisite for modern experimental knapping. The four case-studies that form the body of this thesis are performed with explicit hypotheses and possess different degrees of scope and methodological control, therefore offering an opportunity to test the significance of hypothesis testing and the efficacy of the model outlined in Chapter 2.

The first case study (Chapter 3) seeks to rectify the unreliable method of flake platform area measurement that is typically used when 3D scanning is unfeasible. This approach involves multiplying the calliper measurements of platform width and thickness, making a rectangular approximation of platforms that grossly overestimates their actual size. A new method is put forth that involves taking measurements according to alternative geometric approximations (triangle, rhombus, trapezoid and ellipse). With platform area often serving to estimate reduction intensity (originally demonstrated by Dibble and Whittaker 1981), which in turn informs interpretations about technological organisation, this new method of platform measurement can increase the reliability of these interpretations. In Chapter 4, a new measure of reduction intensity is developed for backed blades, one of the last implement types lacking a reliable reduction intensity metric. This new metric is applied to the site of Boncuklu in Turkey, and serves to reconstruct the raw material consumption and technological choices at the site. More broadly, this reduction intensity metric can aid the analysis of any backed artefacts made on blade blanks. Chapter 5 concerns the question of whether changes in lithic technology over the course of human evolution were accompanied by increases in raw material efficiency. Measurements of sharp edge perimeter per gram of original core offers insights into any trends or transitions in efficiency. With different hominins possessing different tool-kits, these findings could relate to the technological efficiency of past hominins. Lastly, Chapter 6 addresses the enigmatic concept of behavioural complexity by offering means of quantifying hierarchical complexity. This is achieved by reconstructing the number and nature of steps involved in different technologies based on footage of expert knappers. Trends in the extent of hierarchical organisation involved in different technologies serve to inform the behavioural and cognitive abilities of past hominins. Aside from the findings in these individual papers, these diverse case-studies also offer an opportunity to address the following research questions.
1.2 Research Questions
This study seeks to explore the following three main research questions:

1) Which aspects of the archaeological record are informed by knapping experiments?
   While the case-studies in this thesis address flake measurement, reduction intensity, technological efficiency and behavioural complexity, almost all pursuits in lithic analysis can be furthered by experimental knapping. In an attempt to group these pursuits, Eren et al. (2016) identified three key research areas in archaeology that are aided by experimentation; method validation, hypothesis testing, and predictive modelling. The following four case-studies offer steps forward in the former two research areas.

2) How is experimental knapping embedded within the archaeological research process?
   Lithic experimentation is deeply embedded in the research process of archaeology and is associated with other lines of evidence, such as ethnoarchaeology. An aim of this thesis is to reflect on the many ways knapping experiments are entangled within the complex process of archaeological interpretation.

3) How should knapping experiments be conducted and what factors contribute to their validity?
   As mentioned above, a key argument of the model put forth in the following chapter is that the validity of experiments can be judged according to the extent that their scope, control and interpretations align. Emphasis will also be placed on the importance of explicit and falsifiable hypotheses in establishing experiments.

1.3 Thesis organisation
The history of experimental knapping and recent trends in the discipline will be addressed in Chapter 2 in order to develop a model of how the initial parameters of experiments (scope and control) influence the scale of interpretations. The four case-studies in the following four chapters are established with various levels of scope and control, thereby offering an opportunity to test the flexibility and efficacy of this model. Additionally, the use of hypothesis testing in the history of experimental knapping is explored and its significance is again demonstrated using the four case-studies and the explicit hypotheses contained within. Aside from providing a sample with which to test this model, these chapters also advance our understanding of flake measurement (Chapter 3), reduction intensity (Chapter 4), technological efficiency (Chapter 5), and behavioural complexity
(Chapter 6). Finally, Chapter 7 concludes by assessing how well these case-studies conform to the model devised in Chapter 2 and reflecting on the myriad ways that experimental knapping contributes to archaeology and is embedded in the archaeological research process.
CHAPTER 2:
Literature Review and Background

2.1 Introduction
The study of lithic technology has been entwined with experimental knapping studies since the inception of both pursuits. While this thesis explores the role of experimental knapping in developing new methods and testing hypotheses about reduction intensity, technological efficiency and behavioural complexity, there are few, if any, sub-disciplines of lithic analysis that have not benefited from experimentation. The study of reduction sequences, reduction intensity, fracture mechanics, raw materials, heat treatment, decision making, biomechanics, raw material efficiency, skill, behavioural complexity, tool function, projectile technology and taphonomy have all been explored via knapping experiments. Experimental knapping is not without its detractors of course, with some concerned about an unfounded reliance on expert intuition, circular reasoning, limits of parsimony, and the phenomenon of equifinality (Hayashi 1968; Shea 2011; Thomas 1986). In response to these critiques, the last few decades of experimental knapping research involved an increasing reliance on the scientific method, accompanied by explicit and falsifiable hypothesis testing. This chapter explores the history of knapping studies in archaeology to address these, oftentimes valid, critiques and works towards a best-practice model of experimental knapping.

A distinction to be made early in this thesis is between the sometimes conflated definitions of replicative and experimental knapping. For the purposes of this thesis, replicative knapping refers to the breaking apart of non-artefactual stone, or analogous material, to create cores, flakes and tools to in turn inform archaeological interpretations. These replicated lithic products are either intended to as closely as possible resemble specimens from the archaeological record, or to more loosely and cumulatively resemble a particular technology, technique or assemblage. While replicative knapping has a long history in archaeology, experimental knapping is a more recent phenomenon. Experimental knapping is referred to here as any replicative knapping that is embedded within the context of explicit and falsifiable hypothesis testing. Regardless of whether it is conducted free-hand or mechanically, experimental knapping (sensu stricto) is therefore a component of the modern hypothetico-deductive model of scientific reasoning. For this reason, replicative knapping is not a priori experimental. Indeed, most early knapping research involved replicative knapping that was not in the pursuit of explicit hypothesis testing. This became the basis of much resultant criticism which ultimately led to the incorporation of the scientific method to knapping studies.
Archaeological experiments are not solely within the purview of lithic analyses, but have long been recognised as a useful tool for archaeologists more broadly (Ascher 1961b; Coles 1973). These experiments are intended to bridge the gap between the fragmentary archaeological record and the rich assortment of past behaviours we wish to reconstruct. In this way, experimental archaeology shares the goals of ethnoarchaeology. However, where ethnoarchaeology is unsuitable, unfeasible or unreliable, experimentation offers one of the only other avenues for middle-range theory construction.

2.1.1 Experimentation as Middle-Range Theory Construction

The fundamental task faced by archaeologists is reconstructing dynamic and complex past human behaviours using the typically static, fragmentary and taphonomically biased samples afforded by the archaeological record. Perhaps the strongest theoretical scaffolding with which to bridge this static archaeological record with the rich tapestry of human behaviour is middle-range theory. Developed originally in sociology (Larson 1973; Mullins 1973; Parsons 1948; 1950), middle-range theory was later formally adopted in archaeology by Binford (1977b), despite earlier iterations of the underlying principles of the theory being previously foreshadowed (Binford 1968; Binford and Binford 1968). Middle-range theory was originally loosely defined, but can be considered as any theory that links archaeological finds with human behaviour (Binford 1981; Goodyear et al. 1978; Raab and Goodyear 1984). The utility of middle-range theory originally lay in the provision of a theoretical background with which to conduct ethnoarchaeology, most notably demonstrated by Binford (Binford 1977a; 1979; 1980; 1982) with his study of the Nunamiut Eskimo of Alaska and with Australian Indigenous groups. It soon came to, explicitly or not, underpin most experimental studies in archaeology.

An underlying principle of middle-range theory is the concept of analogy, which predates middle-range theory itself. Analogy refers to the use of relevant observable behaviours as a proxy for unobservable behaviours and has been the subject of much discussion in archaeological and anthropological literature (Anderson 1969; Ascher 1961a; Binford 1967; Charlton 1981; Freeman 1968; Gould 1980; Gould and Watson 1982; Hiscock 2008; Hodder 1982; Orme 1974; Stahl 1993; Thompson 1956; Wylie 1985; 1988). While analogy allows an association to be made between two different phenomena, uniformitarianism can extrapolate this link into the past. Uniformitarianism, another fundamental component of middle-range theory, assumes that processes in the past occur in the same way in the present. In terms of lithic technology, the universal and unchanging way in which stone fractures in response to applied force enables lithic experimentation to inform past behaviours.
The use of analogy in archaeology has yielded criticism, especially in the realm of ethnoarchaeology. Some are concerned that much analogical reasoning renders moot the long chronology of the archaeological record, making a static reconstruction of what is typically a dynamic and temporally deep record of past behaviour (Freeman 1968; Gould 1980; Gould and Watson 1982; Hiscock 2008). If applied uncritically with little consideration of the relevance of analogues, then this concern is likely a reality. In response, Wylie (1985) stresses the importance of demonstrating the relevance of analogues to the elements of the archaeological record under investigation. In ethnoarchaeology, relevance of analogical models has historically been argued by considering similarities in geography, environmental niches, subsistence practices or cultural continuity. Meanwhile, verifying relevance in knapping experiments often involves demonstrating the morphological similarities of the archaeological and experimental specimens and debitage, or explicitly defining the archaeologically supported constraints imposed on an experimental knapper. More broadly, the realities of physics and materials science means that stone responds to applied force relatively consistently, thereby limiting the range of variability that can be caused by knapping, past or present.

Temporal depth imposes certain limitations on the use of analogy and middle-range theory construction, as in many cases the principles of uniformitarianism extend only so far. Most pertinent to experimental knapping studies on an evolutionary time-scale, is the millennia of evolution in hand (Marzke 1997; 2013; Rolian et al. 2011) and brain (Frey 2008; Greenfield 1991; Stout and Chaminade 2012; Wynn 2002) morphology that distinguishes the dexterity and decision making of past hominins and present experimental knappers. This concern is most relevant in Chapters 5 and 6, where an attempt is made to quantify the technological efficiency and behavioural complexity of several lithic technologies over the last 3.3 million years. In Chapter 5, two knappers of different skill levels, intermediate and expert, were used to better approximate the varying levels of skill and cognition among past hominins. Additionally, the knapping experiments outlined in Chapters 5 and 6 involved strict adherence to archaeologically, ethnographically and experimentally derived reduction sequences. Adherence to these explicit parameters means that even if the modern expert Homo sapiens knapper found a more cognitively demanding or skilful solution to a problem encountered during knapping, that solution could only be enacted if it was a common component of that technology in the past. These precautions, outlined further in the relevant chapters, serve to strengthen the relevance of the analogies on which those chapters rely. While analogy has been the focus of the valid criticisms discussed here, in lieu of a radically new set of theoretical frameworks, analogical reasoning is the only available means with which to extrapolate meaning into the past.
2.1.2 Experimentation and the Scientific Method

If middle-range theory, analogy and uniformitarianism are the theoretical underpinnings of experimental knapping, then the scientific method forms the methodological foundation. The modern scientific method, involving hypothetico-deductive reasoning, has been applied to knapping experiments with increasing frequency over the last few decades. These studies rely on deductive reasoning and hypothesis testing to gain an understanding of past phenomena. Most simply, deductive reasoning can be understood as a hypothesis-driven approach that progresses logically from generalisations to specificities (Allen 2001; Jones 1909). Alternatively, inductive reasoning is data-driven and progresses from the specific to the general. Inductive reasoning has had its share of prominent detractors (e.g. Hume 1777; Popper 1959; 1963; 1972; Russell 1912; 1945), concerned about the difficulty of demonstrating causality and validity in this paradigm. Within most fields of scientific endeavour, inductive reasoning is seen as the less reliable of the two approaches (Allen 2001; Lawson 2003). Therefore, much research is conducted using the hypothetico-deductive framework.

This model of scientific endeavour also dictates that hypotheses, or predictions of outcomes, be devised according to existing information and then be tested within controlled and observed conditions. Based on this testing, hypotheses are either confirmed or refuted and new hypotheses emerge. Central to this concept is the emphasis on falsifiability, famously espoused by Karl Popper (1959; 1963; 1972), whereby hypotheses only lead to understanding where they are established and tested alongside a means to falsify them, not only validate them. Accordingly, it is argued here that experimental knapping is most reliably conducted with explicitly stated and falsifiable hypothesis testing. As was established in their definitions above, replicative and experimental knapping are seen here as inductive and deductive respectively.

Inductive-based studies are not wholly irrelevant however, and in many fields they play a complementary role with deductive research (Heit and Rotello 2010; Kell and Oliver 2003; Kelley and Scott 2001). In lithic experiments, inductive approaches have served archaeology for decades, and such studies are explored further below. For example, in replicative knapping an expert knapper sets out to recreate a lithic technology with reference to the archaeological record. This approach involves an inductive and post-hoc comparison of the replicated and archaeological lithic assemblages. This method is therefore data-driven and can lead to an understanding of how a certain technology may have been made. While deductive reasoning remains the more theoretically valid of these two scientific approaches, the realities of the natural world sometimes obfuscate a
strict distinction between inductive and deductive reasoning. Due to the uncertainty of what lies under the ground prior to excavation or what lies within a stone prior to knapping, at times a post-hoc and inductive approach is unavoidable in archaeology. In these cases, a data-driven and inductive approach can be the only available option, whereby an archaeologist or knapper selects from a limited range of interpretations to explain the archaeological or experimental record. However, verifying or falsifying this explanation remains difficult with this inductive approach alone. Instead, a replicative understanding of how technologies may have been made can function as the hypotheses within a hypothetico-deductive framework. In other words, the experience and intuitions of expert knappers can contribute to the generation of explicit and falsifiable hypotheses, which in turn offer more theoretically and methodologically robust interpretations.

Accompanying the shift towards hypothetico-deductive science is an increase in the technicality and precision with which experimentation is conducted. This shift has occurred in the natural sciences, as well as in archaeological pursuits like experimental knapping. The central tenets of this experimental approach include replicability, the extent to which an experiment can be conducted repeatedly under the same conditions, and reproducibility, the extent to which an experiment can be conducted under slightly different conditions but culminate in the same conclusion. While replicability reflects the technical precision of an experiment, reproducibility has bearing on the accuracy of experimental findings (Casadevall and Fang 2010; Loscalzo 2012). The so-called ‘reproducibility crisis’ in science is therefore perhaps the most pressing concern for modern scientists (Casadevall and Fang 2010; Cassey and Blackburn 2006; Ioannidis 2012; Loscalzo 2012), especially those in fields that concern human behaviour, such as psychology, economics, social science, and archaeology (Brandt et al. 2014; Cartwright 1991; Earp and Trafimow 2015; Elms 1975; Ferguson and Heene 2012; Francis 2012; Koole and Lakens 2012; Makel et al. 2012; Schmidt 2009; Smith 1970). Methods of tightening replicability and reproducibility are becoming an increasingly prevalent concern in these disciplines. Addressing the systemic issue of a paucity of desire, incentive and outlets for publishing repeated experiments is a logical goal. A more immediately achievable approach is to emphasise the importance of explicit description of methods and the publication of entire data-sets. While concepts like replicability, reproducibility and falsifiability are relatively non-negotiable, albeit sometimes overlooked, aspects of experimentation, elements of experiments that legitimately vary include the scale of the initial scope, level of methodological control, and breadth of the final interpretations.

This chapter seeks to form a best-practice approach to experimental knapping studies based on the initial experimental parameters of scope, control and interpretations. Significant variability exists in
the scale of these parameters among different knapping experiments, the variety of which is explored in the following section. The initial scope of an experiment relates to the spatiotemporal span under investigation and the composition of the sample. For example, experiments with a narrow scope may focus on a particular site or region from a particular time period. An experiment with greater breadth could investigate variation over entire continents and/or over millions of years. Additionally, the sample under investigation can vary, from examining a single aspect of a particular technology to comparing multiple technologies. Meanwhile, the methodological control relates to the extent to which variables are controlled and measured, or free to fluctuate, within the experiment. This is explored in great depth below, especially the contrast between highly-controlled, mechanised experiments and free-hand knapping experiments. Finally, these initial experimental parameters culminate in interpretations based on an experiment’s results, which can vary from the specific to the general. Indeed, all three of these experimental parameters range from specificities to generalities. Exploring the interplay of these three parameters in experiments, and their influence on the validity of an experiment is the key goal of this chapter. To achieve this, the following section considers the history of knapping studies, be they replicative and inductive or experimental and deductive.

2.2 History of Knapping Studies
This section addresses the long and dynamic history of replicative and experimental knapping in archaeology. Other reviews of the history and theory of knapping studies are available (Bradley 1977; Carr and Bradbury 2010; Eren et al. 2016; Flenniken 1984; Lamdin-Whymark 2009; Lerner 2010; Nami 2010; Olausson 2010), the most comprehensive of which addresses the formative years of knapping in lithic technology (Johnson 1978). Apparent in these summaries is that there is an exponential upward trend in the complexity of scientific applications to lithic technology as lithic analysts attempt to keep pace with the emergence of new techniques within and outside the archaeological discipline. Another key trend is the shift from replication experiments designed to comprehend how tools were made, to studies aimed at answering specific research questions via hypothesis testing.

2.2.1 Early Research
Since Nilsson (1868) applied his practical knowledge of gun-flints to prehistoric artefacts and Evans (1860; 1866; 1872) first broke apart a flint nodule to replicate what would come to be known as handaxes, replicative knapping has formed a fundamental component of lithic analyses. From these humble beginnings, replicative knapping saw gradual growth in application and acceptance by early archaeologists during the late 19th and early 20th Centuries, with such studies complementing
much Palaeolithic research (Cushing 1879; Greenwell 1870; Hayes 1890; Holmes 1890; Mason et al. 1891; McGuire 1891; 1892; 1893; 1896; Moir 1912; 1914; 1920; 1926; Pitt-Rivers 1869; 1906; Pond 1930; Sellers 1886; Skertchly 1879; Snyder 1897; Spurrell 1884; Stevens 1870; Warren 1905; 1913; 1914; Wilson 1899). While many of the broader conclusions drawn in these papers would later be overturned and now appear strikingly anachronistic, this was more often a product of the caustic social theories of the period rather than critical fallacies in the underpinning replicative methodology. Key questions arising in this early phase of stone tool replication included the identification of tools shaped by humans, as well as questions surrounding function, flake attributes, stages of manufacture, platform preparation, and differentiating billet types. Much of this information, especially that pertaining to flake attributes and reduction sequences, remains relevant today. For example, the bulb of percussion (Falconer 1868), conchoidal fracture (Evans 1872; Moir 1914), incipient cones (Moir 1920; Skertchly 1879), platform preparation (Spurrell 1884), indirect/punch percussion (Evans 1872) and even fracture mechanics (Pond 1930; Warren 1914) were all identified in these early works.

While the use of replicative knapping in lithic analyses continued to inform archaeological research in the following decades (Clark 1958; Ellis 1957; Goodman 1944; Harner 1956; Knowles 1944; 1953; Tindale and Noone 1941), this pursuit lingered in the periphery of archaeological endeavour until the research and expertise of Don Crabtree (Crabtree 1966; 1967; 1968; 1969; 1970; Crabtree and Butler 1964; Crabtree and Davis 1968; Crabtree and Gould 1970), Jacques Tixier (1963; 1972), and François Bordes (Bordes 1961; 1969; 1970a; 1970b; 1971; Bordes and Crabtree 1969; Bordes et al. 1969) foregrounded the efficacy of replicative and experimental knapping. Key areas of interest in these two decades included reduction sequences (Bradley 1972; 1974; 1975; Crabtree 1966; Newcomer 1971; Rovner 1974; Sollberger 1971), various knapping methods (Bordes and Crabtree 1969; Bordes et al. 1969; Chandler and Ware 1976; Crabtree 1966; 1968; 1970; Henry et al. 1976; Jelinek et al. 1971; Newcomer 1975; Patterson and Sollberger 1976; Peets 1961; Pond 1969; Sheets 1973; Sollberger and Patterson 1976; Stoltman 1971), tool use (Beggerly 1976; Crabtree and Davis 1968), and heat treatment (Bordes 1969; Crabtree and Butler 1964; Flenniken and Garrison 1975; Mandeville 1973; Mandeville and Flenniken 1974; Purdy and Brooks 1971).

2.2.2 Criticisms of These Early Studies

Throughout this long history of replicative knapping studies, there have been those that are broadly supportive of their methodological and interpretative potential and those who critique their very validity and relevance. The concerns of this latter group will be addressed here in turn. While some may immediately dismiss these criticisms as defeatist, many of these concerns are valid and can
contribute to strengthening contemporary lithic experiments. Those critical of replicative knapping have identified an unfounded and heavy reliance on expert intuition, untested acceptance of parsimony, and the problem of equifinality as key problems facing the discipline.

Although a less prevalent concern in recent decades, an over-reliance on intuition based on personal knapping experience has been noted before (Eren et al. 2016; Shea 2011), causing many to question the interpretations gained from early knapping studies. This argument has been most vehemently prosecuted by Thomas (1986:623), who described the ‘flintknapper’s fundamental conceit’ as their belief that ‘the act of breaking rocks gives them the inside track to truth’. Although Thomas (1986) was specifically responding to Flenniken (Flenniken 1984; Flenniken and Raymond 1986), this charge could be levelled at much replicative knapping conducted prior to this time. While the insights of expert knappers are a crucial resource, their intuition is perhaps better directed towards hypothesis generation. Accordingly, an emphasis of the more recent studies considered below is the coupling of an expert knapper’s skill with appropriate, testable and falsifiable hypotheses.

While too much confidence can be bestowed on replicative knappers, the reverse phenomenon can similarly harm the interpretations borne from knapping studies. Underestimating the skill of past and contemporary knappers can take the form of a knapper assuming that their inability to perform a task automatically suggests that no knapper, past or present, could perform such a task. This is best demonstrated by Mewhinney (1964), who challenged his contemporaries to differentiate pressure flakes made with different hammers, following his inability to perform the task. Perhaps unsurprisingly, Muto (1971 in Johnson 1978) later accepted the task and classified the flakes with a relatively high rate of success. Ironically, Mewhinney (1963) had warned against this particular problem only a year prior.

The concept of parsimony has been used to great effect in archaeology. The idea that the simplest or most efficient solution to a problem will most likely be selected by past hominins underpins much behavioural ecology theory in archaeology and optimality models in general (Barlow and Metcalfe 1996; Bright et al. 2002; Broughton 1994; Broughton et al. 2010; Buonasera 2015; Kuhn 1994; Lupo 2007; O'Connell and Allen 2012; Stiner 2001; Stiner and Kuhn 2016; Surovell 2009; Ugan et al. 2003; Winterhalder and Smith 2000). When the idea of optimality is applied to lithic technology, a replicative knapper may discover the easiest, most efficient, or logical means of producing a certain technology. There is no guarantee however, that the same decision making was applied by past hominins. Either the economic and social constraints we infer about the past may not align with those actually imposed on prehistoric knappers, or the most parsimonious solution may not have
been discovered. This problem was highlighted by Hayashi (1968:129) who pointed out that ‘the experimental approach can indicate at best only some possible way(s) in which artifacts could have been manufactured.’ It should be noted that this does not discredit optimality models. Rather, cases where human behaviour deviates from well tested models offers interesting insights into the archaeological record. Instead, experimental knappers should simply consider the limits of parsimony before the interpretative stage of research.

Finally, some are concerned about the issue of equifinality in lithic experiments (Eren et al. 2016; Hayashi 1968; Nami 2010; Shott 1994; Teltser 1991; Thomas 1986), or the phenomena whereby a particular end-product can be achieved via multiple means. Researchers must be guarded against making definitive interpretations where multiple methods of producing the same end-product exist. Analyses of by-products can serve to mitigate the influence of equifinality, as similar end-products produced differently may result in variable patterns of debitage. Additionally, Bradbury and Carr (Bradbury and Carr 1995; 1999; Carr and Bradbury 2000; 2010) recommend lessening the role of equifinality by relying on multiple lines of evidence. In this thesis, the issue of equifinality is most prevalent in Chapter 6, where different technologies were knapped to quantify a component of the behavioural complexity involved in those technologies. As with any task, these technologies can be achieved via multiple possible pathways, making the experiments in Chapter 6 a mere proportion of all possible solutions. To limit the influence of this phenomenon in Chapter 6, the reduction sequences of two knappers were analysed and these knappers conformed to explicitly defined and archaeologically supported constraints on the range of possible knapping techniques.

2.2.3 Addressing Criticisms: Hypothesis Testing and Recent Trends

In response to these largely valid criticisms, the discipline has endeavoured in recent decades to rectify the issues of intuition, parsimony and equifinality by more closely adhering to the scientific method. As discussed above, this primarily involves the hypothetico-deductive model of scientific reasoning, which requires explicit and falsifiable hypothesis testing. Without the cycle of hypothesising, data collecting, interpreting, and hypothesising again, breaking apart rocks offers little to the archaeological discipline.

Along with a more rigorous scientific approach, the methods of experimental knapping have also changed, with a notable trend of increasing technicality with which lithic experiments are conducted. As lithic analyses have grown more complex and precise, so too has the act of knapping and its application to the archaeological record. For example, mathematical modelling is becoming increasingly more complex in experimental lithic studies (Brantingham and Kuhn 2001; Lycett and
Eren 2013a; 2013b). Additionally, digital recording methods like two- and three-dimensional (2D and 3D) photogrammetry have markedly improved our ability to morphometrically compare experimental lithic specimens (Eren et al. 2008; Grosman 2016; Heighway 2011; Muller and Clarkson 2016a; Sumner and Riddle 2008). Finally, the application of 3D scanning to lithic technology and experimentation is a burgeoning technique which has improved our ability to accurately measure lithic attributes (Clarkson 2013; Clarkson and Hiscock 2011; Clarkson et al. 2014; Grosman 2016; Grosman et al. 2014; Lin et al. 2010; Muller and Clarkson 2014; 2016b; Richardson et al. 2014; Sholts et al. 2012; Shott and Trail 2010; Zaidner and Grosman 2015) and make comprehensive analyses of complex technologies (Archer and Braun 2010; Grosman et al. 2011a; Grosman et al. 2011b; Grosman et al. 2008; Li et al. 2015; 2016; Shipton 2016; Shipton and Clarkson 2015a; 2015b). These methods, among others, contribute to increasing the precision and accuracy of lithic experimentation, in part addressing the concerns raised by critics.

2.2.4 Recent Applications of Experimental Knapping
Experimental studies, and lithic experiments in particular, have become such a fundamental component of archaeological research in the last few decades, that to list them all would be a Sisyphean task. Instead, what follows is a summary of how experimental knapping has advanced several sub-fields of lithic analysis. Arguably, the increased relevance of knapping studies to archaeology in recent decades is due in part to the attempts at addressing the concerns raised above and moving the practice towards a more scientifically rigorous approach. Such experiments have influenced interpretations about reduction sequences, reduction intensity, fracture mechanics, raw materials, heat treatment, decision making, biomechanics, raw material efficiency, skill, behavioural complexity, tool function, projectile technology, and taphonomy.

An application of experimental knapping common to both early and more recent research is the identification of features relating to reduction sequences, or the stages and lithic by-products of a particular technology. While experimental knapping has revealed key information about the reduction sequences involved in entire assemblages (Callahan 1987; Reti 2016), these experiments are more often applied to specific technologies. For example, our understanding of bipolar knapping (Kuijt et al. 1995), handaxes (Bradley and Sampson 1986; Shipton et al. 2009; Stout et al. 2014), other assorted bifaces (Amick et al. 1988; Aubry et al. 2008; Shott et al. 2007), Levallois flaking (Tryon et al. 2005), points (Akerman 2007), adzes (Clarkson et al. 2014; 2015b; Shipton et al. 2016) and daggers (Stafford 1998; 2003) have all been influenced by experimental knapping. These studies focus on the successive stages of lithic manufacture and at times highlight diagnostic debitage than can signal the presence of certain technologies even in the absence of end-products.
Instead of analysing variability within technologies, some examine key differences between technologies and their debris (Mauldin and Amick 1989; Odell 1989; Tomka 1989). This poses perhaps the most challenging aspect of lithic variability, but allows lithic analysts to discern different reduction trajectories in commingled debitage assemblages. While this has been attempted with a modicum of success on archaeological specimens (Scerri et al. 2016), Presnyakova et al. (2015) notably differentiated biface debitage from core and flake technologies using an experimentally knapped sample. While some crucial differences in assemblages were identified, the use of calliper measurements in these analyses will likely be further strengthened in the future via more complex 3D scans and analyses.

Variability in lithic technology has also been explored with an emphasis on methods of knapping (Diez-Martin et al. 2011; Li 2016; Madsen 1984; Whittaker 1994), standardisation (Eren and Lycett 2012), use-life (Shott 2002), and reduction intensity in particular. Reduction intensity, or the extent to which stone has been removed from a tool or core, has become one of the primary focusses of lithic analysts as a range of behavioural implications can be drawn from variability in reduction intensity. Much can be discussed regarding the relative merits and limits of the various reduction metrics available, and they are considered in greater detail in Chapters 3 and 4.

The uniformitarian assumption on which all knapping experiments are based is that stone responds to applied force now as it did in the past, due to the way constant laws of physics dictate stone fracture. Many studies have considered the mathematics, physics and materials science involved in the flaking of stone, under the umbrella of ‘fracture mechanics’ (Cotterell and Kamminga 1987; Cotterell et al. 1985; Hayden and Hutchings 1989; Sollberger 1994). Primarily, archaeologists have been concerned with the role of key variables on the resultant morphology of flakes. These variables include platform attributes (Clarkson and Hiscock 2011; Dibble 1997; Muller and Clarkson 2014; Patterson 1981; Pelcin 1997c; 1998), billet type (Driscoll and García-Rojas 2014; Muller and Clarkson 2016a see S1 Text; Pelcin 1997b; Schindler and Koch 2012; Wenban-Smith 1989), strike velocity (Dibble and Pelcin 1995), and core morphology (Pelcin 1997a; Rezek et al. 2011). There are also those who explore the complex interplay among these variables (Dibble and Rezek 2009; Dibble and Whittaker 1981; Magnani et al. 2014).

Other applications of experimental knapping include those more focussed on the stone itself, based on the straightforward uniformitarian assumption that physical properties of stones behave now as they did in the past. With different types of stone being highly spatially varied and often embedded within hunter-gatherer mobility regimes and interaction spheres, raw material types likely influence
patterning of the archaeological record. Accordingly, much experimental research has revealed the role of raw material type on lithic morphology (Archer and Braun 2010; Eren et al. 2011c; Eren et al. 2014; Gurto et al. 2015; Sharon 2008), functionality (Rodríguez-Reñán et al. 2011), and skill (Eren et al. 2011c). In pursuit of method validation, a particularly important role of experimental knapping, some have also studied the role of raw material quality on the reliability of lithic analysis (Driscoll 2011; Tallavaara et al. 2010). While the quality of raw material clearly influences archaeological assemblages, experimental studies have revealed that the quality of some raw materials can be improved via the application of steadily applied and controlled heat. These studies investigate the role of heat treatment on improving the flaking properties of particular raw materials (Brown et al. 2009; Schmidt et al. 2012), and others search for the optimal methods of heat treatment needed to achieve this outcome (Mercieca and Hiscock 2008; Schmidt et al. 2013).

With reconstructing human behaviour a primary goal of archaeologists, more explicit efforts at making behavioural inferences from experimental knapping include those attempting to map the decision making of lithic technology (Burton 1980; Knutsson 1988), and the decision making involved in raw-material selection (Braun et al. 2009; Olausson 1983a; 1983b). A key variable in the decision-making process of stone tool production is the energy and time spent during this process. Several studies consider the time and energy expenditure involved in lithic technologies, such as dagger and axe manufacture (Callahan 1984; Hansen and Madsen 1983; Madsen 1984). Assuming time and energy hold inherent value to past and present knappers, these variables are crucial for comparing the relative costs, benefits and trade-offs involved in economic decisions surrounding lithic technology in the past.

While many of these applications of experimental knapping occur on the tool-type, tool-kit, assemblage, or regional scale, others occur on evolutionary timescales. One of the crucial elements of hominin evolution that facilitated the manufacture of stone tools is the biomechanical evolution of hand and arm morphology required for the prehension and manipulation of cores and hammers. The biomechanics of stone tool manufacture and use have been explored experimentally (Faisal et al. 2010; Rolian et al. 2011; Williams et al. 2012; 2014), with a particular emphasis on the functional advantages of certain hand traits during knapping (Key and Dunmore 2015; Key et al. 2017) and tool-use (Key and Lycett 2011; in press-b).

The raw material efficiency of lithic technologies has also been investigated via experimental knapping. Raw material efficiency refers to the utility of the output of knapping relative to the amount of stone used. It has been compared among a range of different technologies by quantifying
the number and size of usable blanks removed during knapping (Barzilai and Goring-Morris 2013; Diez-Martin et al. 2011; Jennings et al. 2010; Li 2016; Lycett and Eren 2013a; Putt 2015; Rasic and Andrefksy 2001), as well as the length of sharp perimeter produced (Eren et al. 2008; Heighway 2011; Lycett and Eren 2013a; Mackay 2008; Muller and Clarkson 2016a; Prasciunas 2007; Sheets and Muto 1972). These metrics are commonly and most powerfully expressed as a ratio to the mass of the original core or the mass of all blanks removed. The question of whether different technologies perform more efficiently than others is the focus of Chapter 5.

Experimental knapping has also expanded our understanding of the skill involved in knapping throughout the course of the evolution of lithic technology. For example, several studies have assessed the degree of skill required for certain technologies, such as blades (Andrews 2006; Pelegrin 2006), bifaces (Carroll 2016; Nami 2006; Vicky 2005), daggers (Apel 2008; Callahan 2006; Nunn 2006), and Oldowan knapping (Duke and Pargeter 2015; Harlacker 2006). Others have explored the role of skill in determining varying levels of knapping success, commonly quantified by the efficiency of knapping and the incidence of knapping errors (Eren et al. 2011b; Eren et al. 2011c; Muller and Clarkson 2016a; Nichols and Allstadt 1978; Nonaka et al. 2010; Shelley 1990; Stout et al. 2014; Stout and Semaw 2006). Experiments have revealed knappers of varying skill levels in the archaeological record (Carroll 2016; Ferguson 2008; Fischer 1989; Pigeot 1990), and even identified instances of craft specialisation (Arnold 1987; Olausson 1993; 1998) and the mechanics of skill transmission, namely imitation, gestures and verbal cues (Geribás et al. 2010; Putt et al. 2014; Shipton 2010). Finally, experimentation has facilitated attempts at identifying individual knappers and their idiosyncratic skills and behaviours in the archaeological record (Eren et al. 2011b; Gunn 1975; Shelley 1990; Stahl 2008; Whittaker 1987).

A prevalent concern of archaeology and palaeoanthropology is the evolution of hominin behavioural and cognitive complexity. While instances of symbolism and fossil evidence provide trace evidence for these phenomena, these traits preserve poorly and haphazardly in the archaeological record. The significantly more durable and widespread lithic record offers a less taphonomically biased signature for complex behaviour and cognition. A key task for those conducting experimental knapping has been to find components of lithic technology that correlate with aspects of behaviour and cognition (Eren and Lycett 2012; Mahaney 2014; Moore 2010; Pelegrin 1990; Shipton et al. 2013b; Shipton and Nielsen 2015; Shipton 2013; Shipton et al. 2009). Another approach has been to examine relative cognitive traits via scans of brain activity and blood flow (Stout and Chaminade 2007; Stout et al. 2000; Stout et al. 2008; Uomini and Meyer 2013). Chapter 6 offers a step forward in the former approach by exploring the association of lithic
technology with behaviour by quantifying the hierarchical organisation involved in various knapping technologies. Any experiment that seeks to reconstruct the behaviours of past hominins are based on the uniformitarian assumption that commonalities exist in the decision making processes of past and present knappers due to common goals and the limited number of decision pathways that can achieve these goals. As this uniformitarian assumption is not as straightforward as those relating to the inherent properties of stone, precautions must be taken in experiments dealing with decision making of past hominins. Where differences in knapping choices may occur due to the millennia of hand and brain evolution that differentiate past and modern knappers, enforcing adherence to explicit reduction sequences can mitigate the impact of these evolutionary differences. This precaution was fundamental to the behavioural experiment in Chapter 6, where the assumptions, reduction sequences and experimental parameters are described in more detail. These studies also have potential implications for the development of syntactic language, with recent experimental evidence pointing towards a co-evolution of complex-tool making behaviour and language (Mahaney 2014; Morgan et al. 2015). Any complex behaviours observed in Chapter 6 may therefore have potential bearing on the complex behaviours involved in syntactic language.

While a major component of experimental knapping centres on the output of knapping and the knapping itself, many other experiments instead reflect on how stone tools are used by hominins and how they interact with other elements of the archaeological record. Such studies rely on relatively straightforward uniformitarian assumptions about the material and chemical properties of stones and the objects with which they interact. These studies examine traces of residue and use-wear (Braun et al. 2008b; Claud et al. 2015; Collins 2008; Key 2013; Key et al. 2015; Lemorini et al. 2014; Lerner et al. 2007; Macdonald 2014; Miller 2015; Newcomer et al. 1986; Olausson 1983a; Smallwood 2015), the functionality of particular tools (Bar-Yosef et al. 2012; Cheshier and Kelly 2006; Clarkson et al. 2015a; Couch et al. 1999; Eren and Andrews 2013; Eren and Lycett 2016; Eren et al. 2013; Friis-Hansen 1990; Galán and Domínguez-Rodrigo 2014; Hunzicker 2008; Key and Lycett 2014; 2015; in press-a; in press-c; Key et al. 2016; Lipo et al. 2012; Machin et al. 2007; Nigra and Arnold 2013; Pétillon et al. 2011; Pettigrew et al. 2015; Quinn et al. 2008), projectile technologies (Barton and Bergman 1982; Bergman and Newcomer 1983; Brindley and Clarkson 2015; Clarkson 2016; Crombé et al. 2001; Fischer et al. 1984; Huckell 1982; Hutchings 2011; Iovita et al. 2014; 2016; Lombard et al. 2004; Moss and Newcomer 1982; Pargeter 2011; 2013; Rios-Garaizar 2016; Rots 2016; Rots and Plisson 2014; Sano 2009; 2016; Sano et al. 2016; Schoville 2014; Shea et al. 2001; Sisk and Shea 2009; Titmus and Woods 1986; Weitzel et al. 2014; Wilkins and Schoville 2016; Yaroshevich et al. 2010; Yaroshevich et al. 2013), the lethality of these projectile technologies (Frison 1989; Lombard and Pargeter 2008; Odell and Cowan 1986; Pargeter
2007; Salem and Churchill 2016; Waguespack et al. 2009; Wilkins et al. 2014), and the influence of trampling on lithic movement and damage (Andrefsky 2014; Driscoll et al. 2016; Eren et al. 2011a; Eren et al. 2010; Forssman and Pargeter 2014; Jennings 2011; Marwick et al. 2017; Pargeter 2011; Pargeter and Bradfield 2012; Pevny 2012; Schick and Toth 1993; Temple and Sappington 2013).

2.3 Towards a Best-Practice Approach to Experimental Knapping

2.3.1 A Model of Experimental Knapping

The studies synthesised above reveal the diverse range of research questions that can be addressed with experimental knapping. Also apparent is the variety of experimental methods with which to explore archaeological questions. Some of these studies explore the role of a single variable on minute changes to flake morphology, while others may aim to reconstruct sweeping changes in lithic technology over the last 3.3 million years of stone tool production. While these approaches differ markedly in their scope and methodological control, they do not necessarily vary in their validity.

Scope is defined in this thesis as the spatiotemporal span and diversity of sample to which an experiment can bear relevance. An experiment concerning a single site from a particular time period can be considered to possess a narrow scope, while one examining large regions over several millennia possesses a much broader initial scope. Meanwhile, an experiment with a narrow scope might examine one particular tool type, compared with one with a broader scope wherein a diverse assemblage may be under investigation. Methodological control, the other initial experimental parameter within this model, can be defined as the extent to which variables are controlled and measured within an experiment.

It is argued in this chapter, that experimentation offers valid insights to lithic analyses while there is an equivalency in the scale of the scope, methodological control and ensuing interpretations. For example, an experiment which analyses the role of strike velocity on flake size will not result in interpretations about the size of flakes over the course of evolution in lithic technology, as there is a clear discrepancy between the scale of the scope and interpretations. Meanwhile, this study of strike velocity and flake size could not be achieved by haphazard knapping of different raw materials by different knappers of varying skill levels, due to the mismatch between the level of methodological control and scale of interpretations. Instead, these interpretations would better be achieved on the basis of highly controlled experimentation, likely involving a mechanised flaking apparatus or finely detailed measurements of all contributing variables.
If the scale of the scope, methodological control and interpretations are thought of as variables, then we can consider scope to be directly proportional to the interpretations, and methodological control to be inversely proportional to the interpretations. These linear relationships are shown in Figure 2.1a and b, which when combined culminate in the two-dimensional plane shown in Figure 2.1c. This plane represents how the variables of scope, control and interpretations can vary while maintaining congruent levels of each. Theoretically therefore, a well-conceived and valid lithic experiment should exist on this plane. Experiments situated significantly above or below this plane will likely suffer interpretive flaws due to either too little or too much scope or control. The coloured gradient of Figure 2.1c scales along the lower and upper extremities of the plane, with the blue (darker) extremity representing narrow and highly specific experiments and the yellow (lighter) extremity representing broad and general experiments.

Of the experiments contained in this thesis, Chapter 3 is an example of an experiment with relatively narrow scope and high experimental control. In this experiment, only one flake attribute is under investigation and this attribute is measured with callipers under double-blind conditions. While Chapter 3 therefore represents an experiment near the lower extreme of the plane in Figure 2.1c, studies with marginally narrower scope and greater control could include those that employ a mechanised flaking apparatus. At the other extreme of this plane are studies that investigate broad changes throughout human evolution such as the experiment described in Chapter 6. This experiment concerns the behavioural complexity of technologies spanning the last 3.3 million years, and involves modern knappers recreating the decision making of extinct hominins. Therefore, Chapter 6 would be situated near the upper extreme of Figure 2.1c, with an extremely broad scope and relatively low methodological control.

Another factor to note about this model is that while broader interpretations can encompass a greater sweep of evolutionary history, such interpretations may not be upheld in specific scenarios. Meanwhile, much narrower interpretations are likely to yield more detail about a specific phenomenon. Unlike experiments with broader scope and interpretations, such narrow experiments are likely to contain few exceptions. While the issue of a study’s scope is easily addressed by refining or broadening the research questions posed at the beginning of experimentation, the issue of experimental control is a markedly more complex variable.
Figure 2.1. A model representing the characteristics of theoretically valid knapping experiments. Charts 2.1a and 2.1b represent the relationships between the scale of interpretations and the scale of the scope and control respectively, culminating in chart 2.1c.

2.3.2 Designing a Knapping Experiment: Methodological Control

In the natural sciences, experimental hypothesis testing involves strict control of variables so as to avoid undue influence from confounding variables. With archaeology uneasily straddling the natural and social sciences, such strict experimental controls are not always feasible. Time-depth
and the vagaries of human behaviour at times necessitate a slight loosening of experimental control, as some variables are difficult to observe and others remain altogether unmeasurable. Although experimental knapping studies offer more avenues for control and measurement of variables compared with archaeology in general, the level of control in knapping experiments varies. The role that this extent of methodological control has on interpretations has already been explored in the model above. Within this model, the extent of control alone does not dictate the validity of an experiment, rather it is its relationship to the scope and interpretations that influences validity.

Highly controlled flaking experiments have a long history in archaeological research, with Speth (1972; 1974; 1975) and Faulkner (1972; 1973) revealing key aspects of fracture mechanics. More recent work makes use of a mechanised flaking apparatus to tightly control the influence of the variables involved in knapping (Dibble 1997; 1998; Dibble and Pelcin 1995; Dibble and Rezek 2009; Dibble and Whittaker 1981; Magnani et al. 2014; Pelcin 1997a; 1997b; 1997c; 1998; Rezek et al. 2011). Although such controlled experiments have been criticised for their removal from natural knapping conditions (Tsirk 1974), most of these criticisms were levelled specifically at Speth (1972). Additionally, these controlled studies serve an invaluable role in revealing key variables in flake formation and contribute towards the creation of more focussed and detailed archaeological hypotheses. Most other experimental knappers however, opt for a more archaeologically realistic free-hand method of knapping, using their own hands to manipulate billets and cores. All knapping conducted for this thesis was undertaken using this less controlled but more realistic method.

Both highly controlled mechanised experiments and free-hand experiments have their relative uses and drawbacks. For example, mechanised experiments allow for the measurement of variables such as strike velocity, force and angle, variables that are largely lost to the archaeological record and are difficult to observe in free-hand experiments. Sometimes however, it is the complex interplay of variables that contributes to the end-products of knapping and their fundamental attributes. The control and isolation of key variables may unduly diminish the roles of other interacting variables. Accordingly, some controlled experiments have explored the complex interactions of key variables (Dibble and Rezek 2009; Dibble and Whittaker 1981; Magnani et al. 2014). Free-hand knapping allows for the inclusion of all variables, including the complex interaction of the shoulder, elbow and wrist joints in the final delivery of force. The obvious trade-off in such experiments is a reduction in control.
This trade-off in level of control between mechanised and free-hand knapping experiments influences the type of validity an experiment may hold. Internal and external validity are concepts fundamental to the experiments conducted by natural (Eldridge et al. 2008; Godwin et al. 2003) and social (Roe and Just 2009; Schram 2005) scientists alike, with different studies possessing different relative levels of internal and external validity. Internal validity refers to the certainty with which a correlation between the dependent and independent variables within a study can be considered causal. Increasing levels of control, conducting blind experiments, restricting the influence of confounding variables, and increasing sample size can all improve the internal validity of experimentation. Meanwhile, external validity refers to the extent to which the results of a study can be transposed to other experimental settings, or enacted in real-world scenarios. Broadening a sample and making experiments more closely resemble natural conditions can improve this type of validity.

These concepts have primarily been applied to archaeological experimentation by Eren and Lycett (Eren et al. 2016; Lycett and Eren 2013b), who view highly controlled, mechanised knapping experiments as highly internally valid, and more realistic, free-hand knapping experiments as highly externally valid. The four experiments outlined in the following chapters all possess more external than internal validity, as they all involved free-hand knapping. The experiments in Chapters 4, 5 and 6 required the production of particular stone tool typologies, a task not readily achievable using a static and mechanised knapping apparatus. Therefore, a mechanised method may have slightly improved the internal validity of the experiments, but would have sharply reduced the external validity, or the real-world applicability, of these studies. Within this scheme, much like the model posed above, neither approach is intrinsically more valid. Instead, different methodological approaches possess different types of validity, meaning that method selection is best conducted based on the kinds of research questions being posed.

It is clear that neither mechanised nor free-hand knapping methods are superior. Eren et al. (2016) therefore advocate a pragmatic approach to method design. As such, the level of control involved in an experiment should be selected based on the hypotheses under investigation. Again, a congruency in the scale of the scope, methodological control and interpretations should determine the level of control required by an experiment. This does not necessarily require a partitioning of controlled and free-hand experimentation however, as both approaches can play complementary roles in hypothesis creation and testing.
When a free-hand knapping approach is selected, another consideration that influences the extent of methodological control is the experience and skill of the knappers under investigation. A large body of work has clearly demonstrated the influence of knapper skill on the end-products of knapping (Eren et al. 2011b; Eren et al. 2011c; Muller and Clarkson 2016a; Nichols and Allstadt 1978; Nonaka et al. 2010; Shelley 1990; Stout et al. 2014; Stout and Semaw 2006). Some studies (including Chapter 5) opt for a qualitative comparison of knapper skill, typically dividing knappers into ‘experts’, ‘intermediates’ or ‘novices’, or some derivation of those terms. For studies interested in a finer detailed approach, knapper skill can be expressed in quantitative terms, such as hours or years of experience.

When knapping experiments occur on an evolutionary timescale and include technologies made by our hominin ancestors, an inherent shortcoming is our access to only one species of knapper; Homo sapiens. While this may seem an inescapable pitfall of experimentation, several methodological options exist to minimise this limitation. In Chapter 5, an attempt is made to accommodate a wider range of hominin knapping ability by incorporating the results of both an expert and intermediate knapper. Additionally, in both Chapters 5 and 6 the knappers conform to explicitly defined and archaeological supported constraints on their knapping. Another approach is to observe the knapping of extant species of apes, such as the studies performed with Abang the orangutan (Wright 1972) and Kanzi the bonobo (Toth et al. 2006). Like experiments involving Homo sapiens knappers, these studies observe the behaviour of species with common ancestors to the hominins under investigation. While neither approach accurately reconstructs the brain or hand structures of past hominins, by exploring the knapping abilities and products of both Homo sapiens and extant species of apes archaeologists can effectively book-end the evolutionary range of past hominin knapping ability.

Another aspect of experimental control is the extent to which a study is conducted blind. Studies in the natural sciences commonly include blind experimentation, where either the participants are unaware of the study objectives and sample (single blind), or both the participants and experimenters are unaware (double blind). For example, the calliper measurement portion of the experiment in Chapter 3 was conducted under double-blind conditions, where neither the authors nor participants were aware of the results of the more reliable 3D scanned measurements. While such methods are ideal, they are not always feasible in experimental knapping studies, where constraints on sample size and number of available participants can limit the extent to which blind studies can be conducted (Eren et al. 2016). Due to such constraints, the blade blanks in Chapter 4 were knapped by CC and retouched by AM. Although the production of blade blanks was
conducted blind, the retouching was conducted under non-blind conditions as these pieces could only be retouched following the analysis of their blanks. Eren et al. (2016) point out that where blind testing is unfeasible, the problems associated with non-blind studies can be largely mitigated by restricting as much information as possible, as well as by conforming to archaeological templates or to explicitly defined parameters. Accordingly, to limit the influence of the non-blind conditions in Chapter 4, the retouching conformed to the common typological forms of backed blades. This was also the approach adopted in Chapters 5 and 6. In these experiments, the two knappers were aware that the studies related to technological efficiency and cognitive complexity respectively, but at the time of knapping were unaware of the methodology to be used to quantify these phenomena. The knappers were informed of the goals of the study, but not the methods, so that they could tailor their insights provided during and immediately after the experiments. This was especially significant in Chapter 6, where these insights helped construct the hierarchical diagrams that form the basis of the study.

2.3.3 Designing a Knapping Experiment: Sample Size
Related to the issue of methodological control is the nature and size of the composition of an experimental sample. In experimental knapping the sample includes the knappers and their knapped products, and is a consideration that must be established early in a study. Williams and Andrefsky (2011) found that the debitage from five different modern knappers displayed significant morphological variation, highlighting the importance of selecting an appropriate number of knappers in cases where this knapper variation may impose confounding influence on the results. Therefore, where appropriate in this thesis, multiple knappers are used to explore the role of individual knapping variation.

In Chapters 3 and 4, this was not necessary as the output from knapping alone was the subject of investigation, not the process of knapping. In Chapter 3 the platforms of a sample of experimentally produced flakes are analysed, and in Chapter 4 a suite of measurements are taken on blades and backed blades. In both cases, the knapper and their skill level are relatively unimportant, as long as the sample is suitably diverse while existing within the bounds of definitions of flakes, blades and backed blades. Chapters 5 and 6 on the other hand, examine broader trends in the evolution of lithic technology using large experimentally knapped assemblages. In these studies, idiosyncrasies of individual knappers could potentially unduly influence any observed patterns and two knappers are used to explore this potential confounding factor. Therefore, the output of two knappers is analysed in both experiments.
While the inclusion of more knappers will invariably strengthen the results of experiments, inherent time, budgetary and availability constraints are often imposed on the total number of participants involved. As Eren et al. (2016) point out, technologies like Levallois and blade knapping require years of practice, and the even more technical Danish and Egyptian daggers and knives require decades of experience. The availability of people who meet the necessary skill requirements often imposes severe limits on sample size.

2.3.4 Designing a Knapping Experiment: Relevance to the Archaeological Record

A crucial consideration in the design of a knapping experiment is its applicability to archaeology. Knapping experiments should be borne from a founding basis in the archaeological record, and any knapping experiment entirely divorced from reference to archaeology is exposed to easy criticism. While consideration of artefacts is vital, studies which comprise only experimental data may still bear relevance to archaeology. Indeed, three of the four case-studies explored in this thesis do not rely on archaeological data. All case-studies however, rely on the archaeological findings of others to develop hypotheses, select methods, and make interpretations. Additionally, these case-studies could easily be applied to the archaeological record in the future, and each bears mention of potential avenues for applying the findings to archaeological examples. At the very least, hypotheses should be developed with archaeological relevance in mind, with the subsequent interpretations yielding direct consequence to the archaeological record.

In summary, it is recommended in this thesis that the scale of the scope, methodological control and interpretations in knapping experiments be suitably compatible. The following four experimental case-studies offer varying degrees of scope, control and interpretations, and will thereby serve as a test of the model outlined in Figure 2.1. Also highlighted above is the importance of clearly defined and stringently controlled hypotheses based on appropriate prior research, data or theory. Accordingly, the following section outlines the hypotheses to be tested in the following four chapters. These hypotheses additionally serve to test the research questions outlined in Chapter 1.

2.4 Case Studies and Hypothesis Testing

The following four chapters are comprised of four journal articles, each addressing a different aspect of lithic technology via experimental knapping. Eren et al. (2016) define three key research areas that are most frequently served by knapping experiments; hypothesis testing, modelling, and method validation. The case-studies outlined in this thesis satisfy the first and last of these research areas, with Chapters 4, 5 and 6 serving to test hypotheses about the archaeological record, and Chapters 3, 4 and 6 acting to validate new methods of analysis. Specifically, these four case-studies
employ experimental knapping to explore platform measurements, reduction intensity metrics, raw material efficiency, and behavioural complexity respectively.

2.4.1 Hypothesis Development: Chapter 3
The typical method of flake platform area calculation involves multiplying calliper measurements of platform width by platform thickness. This results in a rectangular approximation of platform shape, an approximation that falls far short of the wide range of geometric and amorphous shapes that comprise real flake platforms. 3D scanning technology has markedly improved our ability to model and measure flakes, especially the small and difficult to measure features like platforms. In scenarios where implementation of 3D scanning technology is unfeasible however, due to time, budget, electricity or raw material constraints, lithic analysts must rely on calliper measurements. Chapter 3 seeks to validate a new method of calliper based platform area measurement that involves simply measuring the platforms according to other geometric approximations (triangle, rhombus, trapezoid and ellipse). The testing of this method is conducted on a sample of experimentally knapped flakes.

As this paper is primarily focussed on the development and validation of a new method, little hypothesis testing is involved. However, the underlying null hypothesis is that the newly developed method of platform area measurement will be no more accurate or precise than the existing method of multiplying platform width and thickness. If this hypothesis is rejected however, the new method of platform measurement could serve analysts, when 3D scanning is unfeasible, who base reconstructions of tool reduction intensity on platform size. In turn, reduction intensity can inform a range of interpretations about technological organisation and human behaviour.

2.4.2 Hypothesis Development: Chapter 4
This critical role of reduction intensity metrics in reconstructing aspects of human behaviour has resulted in decades of research devoted to developing and testing new methods of estimating reduction intensity. As such, very few tool and core types currently lack a suitable specific or universal reduction metric. Backed blades are one of these few remaining tool types lacking such a metric. The first component of this paper will develop and test the efficacy of a new allometric reduction measure for backed blades on an experimentally knapped sample. This new reduction intensity metric allows for a reconstruction of the size of the blanks from which backed blades are made. Based on the original size of backed blades from Boncuklu, a model of raw material and blank consumption can be developed. The second component of this paper will test the hypothesis that the production of backed blades at Boncuklu contributed to preserving their limited supply of...
raw material and that the modular functionality of backed blades enabled the inhabitants to navigate the transition from the Epipalaeolithic to the Neolithic.

This hypothesis is based on the existing literature which views backed blades as a remarkably efficient, adaptable and maintainable technology (Clarkson et al. in press; Hiscock 1994; 2002; Hiscock et al. 2011; Neeley 2002). Additionally, the scarcity of raw material at the site, due in part to the distance from sources, would likely either have necessitated or be alleviated by strategies that efficiently consume raw material. As backed blades were common at various times during the Middle Stone Age through to the Neolithic in Europe, the Near East, Africa, South Asia, and parts of Australia, this new reduction intensity metric could have applications far beyond the Neolithic of Turkey.

2.4.3 Hypothesis Development: Chapter 5
The evolution of lithic technology has long been thought to be accompanied by broad improvements in raw material efficiency. This paper seeks to test the raw material efficiency of eight different lithic technologies, broadly spanning the Oldowan to the Neolithic. Raw material efficiency is here estimated using photogrammetry to calculate the length of sharp flake perimeter per gram of the original core. These eight different technologies were experimentally knapped by an expert and intermediate knapper.

This experimentally knapped sample is used to test the null hypothesis that no significant differences in sharp edge length per gram exist among the eight different lithic technologies. This hypothesis is based on previous studies which found equivalencies in raw material efficiency between multiplatform, discoidal, bifacial and blade cores (Eren et al. 2008; Jennings et al. 2010; Prasciunas 2007; Rasic and Andrefksy 2001). This study was instigated following concerns that the approach of previous studies to compare only two different technologies may be overlooking differences in a broader range of lithic technologies. If this null hypothesis is overturned, broad trends and key transitions in raw material efficiency may be identified, potentially contributing to our understanding of explanations for technological change and variability.

2.4.4 Hypothesis Development: Chapter 6
As with the previous chapter, this paper explores sweeping trends in the evolution of lithic technology. Specifically, quantifications of the hierarchical organisation of five different lithic technologies are used to estimate the behavioural complexity required for their manufacture. Two
expert knappers were filmed, producing different iterations of these five technologies and this footage was subsequently analysed to reconstruct the number and nature of the steps involved.

This paper tests the hypothesis that differences exist in the behavioural flexibility required for knapping different technologies. This hypothesis is based on the results of previous studies as well as a swathe of assumptions and hypotheses currently untested in the literature. This is expressed as a series of hypotheses outlined in the Hypothesis Testing section (6.1.1). As with the previous case-study, the null hypothesis is that no differences exist among the various technologies under investigation. If this null hypothesis is rejected, these results could have implications for the behavioural and cognitive abilities of past hominins.

2.5 Summary
These four case studies employ varying degrees of scope and methodological control that lead to varying breadths of interpretation. They therefore offer an opportunity to test the model outlined in Figure 2.1. The validity of these four knapping experiments can be weighed according to the extent that the scope, control and interpretations are aligned. The set of explicit and falsifiable hypotheses outlined above will serve to demonstrate both the importance of the hypothetico-deductive model of scientific reasoning, and examples of possible archaeological applications potentially served by experimental knapping.
CHAPTER 3:
A new method for accurately and precisely measuring flake platform area


**Abstract**

The use of 3D scanning to measure platform size is remarkably accurate and precise and has been used to estimate reduction intensity in lithic assemblages. While using 3D scans to measure lithic artefacts remains best practice in archaeology, the use of 3D scanning technology is seldom possible in real-world archaeological scenarios. Time constraints, budgetary constraints, large sample sizes, remote fieldwork locations, unreliable access to electricity and unsuitable raw material types can render 3D scanning unfeasible. In such scenarios, platform size is typically estimated by multiplying calliper measurements of platform width and thickness, forming a rectangular approximation of the platform. As few platforms resemble rectangles, we find that this method of platform measurement approximately doubles true platform size and is neither accurate nor precise. To remedy this overestimation, this paper introduces and tests the Geometric Approximations of Platforms (GAP) method, where simple 2D shapes (triangle, rhombus, trapezoid and ellipse) are employed in favour of the currently used rectangular approximations. Compared with platform width and thickness measurements, the GAP method is an equally simple and easily applied method that significantly increases both the accuracy and precision of platform measurements, being statistically indistinguishable from 3D scanned platform area values. Moreover, we offer a means by which platform width and thickness values can be converted to be used in conjunction with those obtained from the GAP method, meaning that inter-site comparisons with previously analysed assemblages would not be hindered for those employing the GAP method.

**3.1 Introduction**

The importance of flake platform attributes has long been understood, with platform area an almost universally measured variable in global lithic analyses. This present study aims to refine the accuracy and precision of platform area measurements in the interests of improving the existing behavioural models that are reliant on platform measurements. In particular, much attention has been given to the relationship between platform variables and flake size as a means of estimating reduction intensity. For decades, estimating the extent of reduction or curation of lithic artefacts has been a key concern in archaeology as these features can be used to model technological

The relationship between platform variables and original flake size was originally exploited by Dibble and Whittaker (1981), and much subsequent attention has been directed at confirming and refining our understanding of this relationship (Braun et al. 2008a; Clarkson and Hiscock 2011; Davis and Shea 1998; Dibble 1995; 1997; 1998; Dibble and Pelcin 1995; Dibble and Rezek 2009; Dogandžić et al. 2015; Lin et al. 2013; Magnani et al. 2014; Muller and Clarkson 2014; Pelcin 1997a; 1997b; 1997c; 1998; Rezek et al. 2011; Shott et al. 2000). By creating a linear regression between platform measurements and original flake mass, predictive equations can be developed to estimate the original mass of flakes. Reduction intensity can then be measured by comparing this predicted original mass with the actual mass of the retouched artefact.

Many of the early reduction measures based on regressions of platform attributes and flake mass suffered from low coefficients of determination, inhibiting the predictive power of these measures. At the time, Dibble (1998) pointed out that a ratio of original to predicted mass was not an effective measure of reduction, as the error involved in the predictor equations was too high to account for finer retouch. This was largely due to the use of callipers to manually measure platform width and thickness, with the result of multiplying these values being used as the estimate for platform area. This is the standard method of measuring platform area globally. In effect, multiplying platform width and thickness measurements approximates all platforms as perfect rectangles. Lithic analysts will know however, that few platforms even roughly resemble rectangles.

Dibble (1997; 1998) has long identified this source of measurement error, as rectangular approximations are ill-suited to the majority of actual platforms which are markedly variable in shape and at times amorphous. Studies have shown that rectangular approximations of platforms result in platform area measurements that significantly overestimate actual platform area (Braun et al. 2008a; Clarkson and Hiscock 2011). Recently, accurate and precise measurements of platform area using 3D scanning have enabled marked improvements in the power of regressions used to infer original flake size from platform area (Braun et al. 2008a; Clarkson and Hiscock 2011; Muller and Clarkson 2014).

Beyond estimating reduction intensity, 3D scanning approaches to lithic analysis are becoming more common, and powerful conclusions have been drawn using such technology including
morphology (Bretzke and Conard 2012), reduction sequences (Clarkson et al. 2014; 2015b; Goren-Inbar et al. 2011), typology (Grosman et al. 2008), cortex ratios (Lin et al. 2010) and taphonomy (Grosman et al. 2011b). However, such 3D scanning approaches remain costly and time consuming, thus measuring platform area with 3D scans is rarely feasible, especially in field archaeology scenarios. As such, the majority of global lithic analyses still employ platform width and thickness measurements for platform area.

A principal reason 3D scanning may by unfeasible is the time constraints involved with many archaeological projects, potentially exacerbated by large sample sizes or scenarios where analysis must be conducted during the field season. Additionally, with much fieldwork being conducted in remote locations, a lack of access or unreliable access to electricity can hinder in-field 3D scanning. While commercial entry-level 3D scanners are becoming increasingly affordable, many are still prohibitively expensive. Finally, particularly fine-grained and glossy raw materials, such as obsidian, are poorly suited to 3D scanning. While industrial adhesives can be applied to the surface, in the interests of use-wear and residue studies this is rarely desirable. This present research was borne out of the authors’ analysis of stone tools at Boncuklu, an early Neolithic site in Turkey (Baird et al. 2012). The most ubiquitous raw material is obsidian, flaked stone is prohibited from being exported, field season duration is strictly limited, and black-outs and brown-outs are common. Although this is only one example where 3D scanning is impractical, any of the factors described here could be prohibitive to 3D scanning.

For these reasons, it is necessary to develop a quick, inexpensive and easy means of measuring platform area that is not encumbered by low accuracy and precision. This study attempts to remedy the poor accuracy and precision of conventional calliper measurements that approximate platform area with a rectangle by introducing and testing the Geometric Approximations of Platforms (GAP) method, where simple 2D shapes (triangle, rhombus, trapezoid and ellipse) are employed in favour of the currently used rectangular approximations.

3.2 Materials and Methods
To test the efficacy of using 2D shapes such as a triangle, rhombus, trapezoid and ellipse, over the conventional rectangular approximation of platform area, a sample of 123 experimentally knapped flint flakes with intact platforms was used in all analyses. The first component of the analysis involved measuring all platforms with the existing convention of platform width and thickness, and then with each of the different geometric approximations. These digital calliper measurements (to
the nearest 0.01mm) were conducted according to the definitions and equations of the 2D shapes outlined in Table 3.1.

The geometric approximations were constructed in the analyst’s mind while measuring, with the aim of forming an approximation as close as possible to the actual platform, while not violating the geometric equations (See Figure 3.1 for examples). The four new 2D shapes can be imagined in any orientation, and are not bound by any flake attribute as existing rectangular approximations are by the axis of flaking. For the existing rectangular method of measuring platform area, platform width (x) was measured arbitrarily perpendicular with the percussion axis, and platform thickness (y) measured perpendicular to the width. The inflexibility of this convention does not adequately account for the variability in platform morphology. Platform morphology varies markedly according to different knapping techniques, percussor types, striking angles and core morphologies. As such, very few archaeological or experimental platforms closely resemble rectangles.

The method proposed here introduces new geometric approximations. The triangle equation is applicable for all triangle types including equilateral, isosceles, scalene and right-angled triangles. This approximation is archaeologically most common on flakes possessing triangular cross-sections (i.e. one arris or prominent dorsal scar intersection). Meanwhile, the rhombus approximation is typically best suited to dihedral platforms, the trapezoidal approximations are often suited to flakes with trapezoidal cross-sections, and the ellipse approximations are especially useful for focalised platforms.
Table 3.1. Descriptions of each of the different geometric approximations used in this study, accompanied by the equations used to calculate platform area.

<table>
<thead>
<tr>
<th>Geometric Approximation</th>
<th>Definition</th>
<th>Equation (Area=)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rectangle</strong></td>
<td>Four sided polygon with four right angles. In lithic analyses, x and y are platform width and thickness. Very few platforms resemble rectangles.</td>
<td>$x \times y$</td>
</tr>
<tr>
<td><strong>Triangle</strong></td>
<td>Three sided polygon, where y is the perpendicular height from the tip to the base (x) of the triangle.</td>
<td>$\frac{x \times y}{2}$</td>
</tr>
<tr>
<td><strong>Rhombus</strong></td>
<td>A parallelogram with four equal sides, where x and y are diagonals between the tips of the rhombus which meet perpendicularly.</td>
<td>$\frac{x \times y}{2}$</td>
</tr>
<tr>
<td><strong>Trapezoid</strong></td>
<td>An isosceles trapezoid is a four sided polygon of z height, with a pair of parallel sides (x and y) also possessing equal base angles.</td>
<td>$\frac{x + y}{2} \times z$</td>
</tr>
<tr>
<td><strong>Ellipse</strong></td>
<td>An oval shape, where x and y are the length of the smallest and largest diameters of the ellipse respectively.</td>
<td>$\pi \frac{x \times y}{4}$</td>
</tr>
</tbody>
</table>

These four geometric shapes put forward as part of the GAP method were chosen based on these archaeological correlates to best summarise a wide variety of platform shapes, while maintaining methodological simplicity and efficiency. All but one of these 2D shapes requires only two calliper measurements, the same number currently used in rectangular approximations. The trapezoid approximation requires three measurements however, representing only a minimal increase in measurement time. Simple equation fields in existing lithic recording databases could easily
accommodate this methodological change. More complex 2D shapes were excluded from this method in the interests of efficiency. If another platform shape is common at a particular site, then this method can easily be modified to incorporate that 2D shape. For example, in assemblages with many dihedral, faceted or chapeau de gendarme platforms, summing the areas of two or more 2D shapes could be used to accommodate the convex and multi-surface morphology of these platforms. In the interest of simplicity, this modification was not considered here, but no added imprecision or inaccuracy would be expected.

![Figure 3.1. Photographs of platforms superimposed with the four new geometric approximations. Platforms were painted red in the second portion of the experiment to facilitate more accurate 3D scan measurements.](image)

Following the calliper measurements of all geometric approximations, 3D scans were conducted on the platforms to highly accurately measure platform area (±0.005” or 0.127mm). Platforms were painted red to make identification of the platforms in the scanned images more reliable. Scans were obtained using a NextEngine HD 3D scanner and the resultant images were trimmed and their area measured using the provided Scan Studio software.

As a quantitative test of the GAP method, each calliper based platform area calculation was expressed as a ratio in terms of the 3D measurements. Due to the remarkable accuracy and precision of scanned 3D platform measurements, we assume these to be ‘true’ values of platform area. Accordingly, ratios of approximately 1.0 can be considered to be very close to the true platform size, whereas ratios of 2.0 are double the true platform size. All 2D geometric approximations were applied to each platform, generating five platform area values (rectangle, triangle, rhombus, trapezoid and ellipse). As only one approximation can be used in real lithic analyses, a prediction
was made for each flake regarding which geometric approximation would best suit the platform shape. This prediction was made before any measurements were taken in order to avoid any influence on decision making from knowing the area measurements.

The ratio to the 3D platform area values for the conventional rectangular measurements was plotted against this ratio for two different sets of GAP measurements. The first, called the ‘actual closest approximations’ is comprised of results in which all selections of geometric approximations were closest to the ‘true’ measurement (i.e. a ratio of 1.0). In this way, the set of actual closest approximations represents the ‘best-case’ scenario, with the best 2D approximation being selected for all 123 platforms. As would be the case in real lithic analyses, there is likely to be some incorrect selections of which geometric approximations would most closely resemble the actual platform. For this reason, the second set of GAP measurements is called the ‘predicted closest approximations’ and reflects the 123 platform area values that were selected before measurement based on the analyst’s prediction of which 2D shape best reflects the platform shape.

One-sample t-tests with Bonferroni-adjusted p-values (counteracting the increased risk of a type-I error during multiple comparisons) were used to test the accuracy (closeness to a known value) of both sets of GAP measurements, with a ratio value of 1.0 used as the population mean. Meanwhile, the precision of the rectangle approximation values was compared with the GAP measurements using Levene’s test for equality of variances. As precision is statistically the inverse of variance (i.e. $1/\sigma^2$) a high variance indicates low precision and a low variance indicates high precision. For the GAP method to be worthwhile, it needs to be both accurate and precise.

### 3.3 Results

The measurement and ratio data in Table 3.2 demonstrate that rectangular measurements drastically overestimate the actual platform area, confirming the findings of previous studies (Braun et al. 2008a; Clarkson and Hiscock 2011). On average, rectangular approximations in this sample overestimate the 3D platform measurements by 19.97mm$^2$.

#### 3.3.1 Accuracy

The overestimation of rectangular measurements is also manifest in Figure 3.2, which compares these measurements to the actual closest and predicted closest sets of measurements. As a test of accuracy, we use a ratio value of 1.0 as the ‘true’ platform size, reflected by the horizontal line in Figures 3.2-4. A series of one-sample t-tests with Bonferroni-adjusted p-values demonstrate that the mean of the rectangular approximations significantly deviates from the ‘true’ platform size ($t =$
30.64, p < 0.001). Meanwhile, the actual (t = -0.84, p > 0.99) and predicted (t = -0.70, p > 0.99) closest approximations are not significantly different than the given value of 1.0. Therefore, while rectangular approximations, derived from platform width and thickness measurements, statistically overestimate true platform size, both the actual closest approximations and the predicted closest approximations are not statistically different than the ‘true’ platform size measurements based on 3D scans. With both the actual closest (‘best-case’ scenario) and predicted closest (‘real-world’ scenario) sets of platform measurements being statistically indistinguishable from the extremely accurate 3D scanning measurements, we can conclude that the GAP method offers a highly accurate means of measuring platform area, with significant advantages in accuracy over the conventional rectangle approximation method.

Table 3.2. Summary data (N = 123) for all different measurement sets, showing mean and spread results for measurements and values for the ratio to 3D measurements.

<table>
<thead>
<tr>
<th>Measurement (mm²)</th>
<th>Mean (µ)</th>
<th>Variance (σ²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D measurement</td>
<td>23.67</td>
<td>1840.94</td>
</tr>
<tr>
<td>All Rectangle</td>
<td>43.64</td>
<td>5603.17</td>
</tr>
<tr>
<td>All Triangle</td>
<td>20.86</td>
<td>1322.96</td>
</tr>
<tr>
<td>All Rhombus</td>
<td>20.03</td>
<td>1301.99</td>
</tr>
<tr>
<td>All Trapezoid</td>
<td>23.46</td>
<td>1851.97</td>
</tr>
<tr>
<td>All Ellipse</td>
<td>27.20</td>
<td>2383.16</td>
</tr>
<tr>
<td>Actual Closest</td>
<td>23.51</td>
<td>1855.12</td>
</tr>
<tr>
<td>Predicted Closest</td>
<td>23.30</td>
<td>1849.96</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ratio to 3D Value</th>
<th>Mean (µ)</th>
<th>Variance (σ²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D measurement</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>All Rectangle</td>
<td>1.98</td>
<td>0.13</td>
</tr>
<tr>
<td>All Triangle</td>
<td>0.93</td>
<td>0.020</td>
</tr>
<tr>
<td>All Rhombus</td>
<td>0.86</td>
<td>0.038</td>
</tr>
<tr>
<td>All Trapezoid</td>
<td>1.01</td>
<td>0.018</td>
</tr>
<tr>
<td>All Ellipse</td>
<td>1.20</td>
<td>0.052</td>
</tr>
<tr>
<td>Actual Closest</td>
<td>0.99</td>
<td>0.0057</td>
</tr>
<tr>
<td>Predicted Closest</td>
<td>0.99</td>
<td>0.023</td>
</tr>
</tbody>
</table>

3.3.2 Precision

Having demonstrated the accuracy of the GAP method, we turn now to an assessment of the precision of this method. As precision is simply the inverse of variance the precision of the GAP method can be estimated using measures of statistical dispersion. A cursory examination of Figure 3.2 reveals that the spread of the rectangle approximation data is exceedingly larger than the actual or predicted closest geometric approximations. This observation is supported by a variety of measures of statistical dispersion, including range, interquartile range (IQR), standard deviation (σ) and variance (σ²) (Table 3.3). According to Levene’s test for equality of variances, the variance of the rectangular approximations is significantly greater than the actual closest (F = 71.39, p < 0.001) and predicted closest (F = 47.50, p < 0.001) approximations. Importantly also, there is no
significant difference between the variances of the actual and closest approximations ($F = 2.40, p = 0.12$). Therefore, there is no significant loss of precision due to the small number of incorrect predictions of which platform would most closely measure platform size. Based on this comparison of variances, we conclude that the GAP method is significantly more precise than the currently used platform area calculation relying on platform width and thickness.

Figure 3.2. Boxplot comparing the rectangular approximations ($N = 123, \mu = 1.98, \sigma^2 = 0.13$) with the set of actual closest geometric approximations ($N = 123, \mu = 0.99, \sigma^2 = 0.0057$), and the set of predicted closest geometric approximations ($N = 123, \mu = 0.99, \sigma^2 = 0.023$). The actual closest approximation set of measurements are comprised of 41 triangle, 25 rhombus, 39 trapezoid and 18 ellipse approximations.

Table 3.3. Measures of spread among the different measurement types.

<table>
<thead>
<tr>
<th>Measurement Type</th>
<th>Range</th>
<th>IQR</th>
<th>Σ</th>
<th>$\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rectangular</td>
<td>2.40</td>
<td>0.34</td>
<td>0.36</td>
<td>0.13</td>
</tr>
<tr>
<td>Actual Closest</td>
<td>0.74</td>
<td>0.064</td>
<td>0.075</td>
<td>0.006</td>
</tr>
<tr>
<td>Predicted Closest</td>
<td>1.73</td>
<td>0.078</td>
<td>0.15</td>
<td>0.023</td>
</tr>
</tbody>
</table>

3.4 The Influence of Selecting the Best Geometric Approximation

The actual closest approximations shown in Figure 3.2 above represent the ‘best-case’ scenario, in which all selections of geometric approximations (i.e. triangle, rhombus, trapezoid and ellipse) were closest to the ‘true’ measurement. As occurred in this experimental application of the GAP method however, some predictions of which geometric approximations would most closely resemble the actual platform were incorrect (17 out of 123). In archaeological analyses, these are not predictions,
but rather selections of which geometric approximation is to be used. In a scenario where 3D
scanning is unfeasible, there is no possibility for an analyst to check how closely their chosen
approximation is to the actual platform size. As predicting which geometric approximation best
suits an actual platform is a subjective endeavour, especially for platforms with particularly
amorphous boundaries, it is necessary to test how important selecting the correct approximation is
to the validity of this method. To do this, calliper measurements were recorded for all geometric
approximations of each platform. A two-sample t-test shows that there is no significant difference
between the actual and predicted closest approximations (t = 0.25, p = 0.80), demonstrating that for
this iteration of the GAP method, there was no significant loss of accuracy owing to the 17
instances where the actual closest geometric approximation was different to the predicted
approximation. This is unsurprising, given that only 17, or 14% of predictions of which
approximation would be most suitable were incorrect.

It is probable that similarly low proportions of incorrect predictions would occur in analyses of
archaeological assemblages. It is therefore likely that there would be no statistically significant loss
of accuracy due to discrepancies in predicted and actually closest geometric approximations, as was
the case in this experimental example. However, for assemblages with higher proportions of
amorphous platforms, which make selection of the most suitable approximation more difficult, it is
possible that the proportion of incorrect predictions might be higher. In order to test a scenario in
which the selection of geometric approximations is less accurate, the rectangle approximation
measurements were plotted alongside the platform measurements of all other geometric
approximations, thereby including the entire range of errors present in non-rectangle
approximations (Figure 3.3). In effect, this is the ‘worst-case’ scenario, as opposed to the ‘best-
case’ scenario (actual closest approximations) outlined in Figure 3.2 above.

One-sample t-tests with Bonferroni-adjusted p-values reveal that while the rectangular
approximations of platform area significantly deviate from the given mean of 1.0 (t = 30.64, p <
0.001), all other approximations remain accurate when compared with the 3D scanned platform
measurements (t = -0.27, p > 0.99). Therefore, even when the entire range of possible errors are
considered and no effort is made to select which 2D shape best approximates the actual platform,
the GAP method still accurately measures platform area.
Figure 3.3. Boxplot comparing the rectangular approximations (N = 123, $\mu = 1.98$, $\sigma = 0.36$, $\sigma^2 = 0.13$) with all other geometric approximations (N = 492, $\mu = 1.00$, $\sigma = 0.22$, $\sigma^2 = 0.049$).

In terms of precision, a Levene’s test for equality of variances reveals that the variance of the rectangular approximations is significantly greater than that of all other geometric approximations combined ($F = 23.17$, $p < 0.001$). Therefore, even when no attempt is made to predict which geometric approximations best suit the platforms, the GAP method remains significantly more precise than the rectangular approximation method. In an archaeological analysis, the chosen 2D approximations are not likely to resemble this ‘worst-case’ scenario, as even a cursory consideration of platform morphology would allow an analyst to relatively reliably choose the most suitable approximation. This is because the four 2D shapes outlined here were chosen based on archaeological correlates to best summarise the variety of platform morphologies typically observed in assemblages. For example, many platforms closely resemble a triangle, rhombus, trapezoid or ellipse, as opposed to the exceedingly few archaeological or experimental platforms which even passingly resemble a rectangle. Having demonstrated that the GAP method is both significantly more accurate and precise than platform width and thickness measurements, even when no attempt is made to select the most suitable 2D shape, we now turn to practical considerations of applying the GAP method.

3.5 Convertibility

Inter-site comparisons are a mainstay of archaeological research. Therefore, some may argue that no degree of increase in accuracy or precision is worth losing the ability to compare assemblages.
For this reason, we offer and test a means of simply and powerfully converting previous platform measurements estimated from rectangular approximations to values that can be used in conjunction with those obtained from the GAP method.

![Boxplot comparing the rectangular approximations (N = 123, μ = 1.98, σ = 0.36, σ² = 0.13) with the halved rectangular approximations (N = 123, μ = 0.99, σ = 0.18, σ² = 0.032), actual closest geometric approximations (N = 123, μ = 0.99, σ = 0.075, σ² = 0.006), and the predicted closest geometric approximations (N = 123, μ = 0.99, σ = 0.15, σ² = 0.023).](image)

**Figure 3.4.** Boxplot comparing the rectangular approximations (N = 123, μ = 1.98, σ = 0.36, σ² = 0.13) with the halved rectangular approximations (N = 123, μ = 0.99, σ = 0.18, σ² = 0.032), actual closest geometric approximations (N = 123, μ = 0.99, σ = 0.075, σ² = 0.006), and the predicted closest geometric approximations (N = 123, μ = 0.99, σ = 0.15, σ² = 0.023).

This conversion involves simply dividing all values derived from a rectangular approximation by the average ratio that the rectangular approximations overestimate the true value, which in this instance is 1.98. In the interests of simplicity however, this value will be rounded to 2.0. Figure 3.4 shows the result of this conversion, with rectangular approximation measurements halved, alongside the original rectangular measurements for comparison.

Testing the accuracy of this conversion, a one-sample t-test reveals that there is no significant difference between these halved rectangular approximation values and the ‘true’ platform area ratio of 1.0 (t = -0.91, p = 0.37). As variance is scale-dependent, it would be expected that the converted rectangular values would be more precise. To test the significance of this change, a Levene’s test for equality of variances shows that the converted rectangular distribution has a significantly smaller variance (F = 22.99, p < 0.001) than the unmodified rectangle approximation values, and is therefore significantly more precise.
Being more accurate and precise than the conventional platform width and thickness method of platform measurement, some may argue that simply halving all platform measurements might be sufficient, rather than using the GAP method. However, when compared to the actual closest (F = 36.18, p<0.001) and predicted closest (F = 11.17, p = 0.001) approximations, the converted rectangular values had a significantly greater variance. Therefore, while halving the rectangular method values is sufficient for inter-site comparisons, there are still easily made gains in precision by measuring platforms with the new 2D geometries.

With improvements in accuracy that make converted platform width and thickness measurements statistically indistinguishable from 3D scanned platform area calculations, and significant improvements in precision, the GAP method not only does not restrict inter-assemblage comparisons, but makes them more accurate and precise. Where platform area is used to estimate reduction intensity for example, these estimations will be significantly more robust if the platforms have been measured using the GAP method.

3.6 Inter-Observer Variability

The final test of the efficacy of the GAP method relates to inter-observer variability, or the degree of variation in results among different users of the GAP method. Platform width and thickness are common and well-known variables in lithic analysis. While the inflexibility of platform width and thickness measurements is partly to blame for the lack of accuracy and precision of this method, it might also afford this method less inter-observer variability. For this reason it is necessary to test the relative inter-observer variabilities of the five different geometric approximations.

Six observers, an author (AM) and five first-time users of the GAP method, participated in a blind experiment in which they measured the platform area of the flakes with all five 2D approximations using the same pair of digital callipers. The five student participants possessed novice to intermediate lithic analysis knowledge, meaning that if the GAP method introduces no added inter-observer variability, it could be safely applied to archaeological assembles even with student analysts. By using novice and intermediate analysts, the following results serve as a baseline for the inter-observer variability of the GAP method. Should this method be employed by expert analysts, its inter-observer variability could only be improved. The student participants had no knowledge of the 3D scanned platform area values or the measurements taken by AM, and were given only Table 3.1 as a guide. As was outlined in the materials and methods section, they were instructed to construct each geometric approximation in their mind with the aim of forming an approximation as close as possible to the actual platform, while not violating the geometric equations. All five 2D
approximations were applied to all 123 flakes, generating five platform area measurements per flake per user, even when a particular 2D shape was clearly not the optimal approximation of the actual platform. This was done in order to test the inter-observer variability even when no effort was made to select the best approximation.

Most archaeological considerations of inter-observer variability focus on the measurement (Clarkson 2002; Dibble and Bernard 1980; Fish 1978; Gnaden and Holdaway 2000; Lyman and VanPool 2009; Wilmsen and Roberts 1978), use-wear analysis (McGuire et al. 1982; Newcomer et al. 1986; Young and Bamforth 1990) or typology (Adams and Adams 1991; Beck and Jones 1989; Boyd 1987; Fish 1978; Whittaker et al. 1998) of materials such as faunal remains, ceramics and stone tools. Many of these studies consider categorical data, and since Fish (1978) first examined discrepancies between analysts in archaeological measurement, there has been little done to standardise an approach to inter-observer variability. For these reasons, two different measures of inter-observer variability are here examined in order to test the applicability of the GAP method.

The first method involves plotting all six individual platform measurements for each flake against the mean value of those six measurements. The merits of this approach have long been known in archaeology, with Dibble and Bernard (1980) using regression between mean values and the individually measured values to test inter-observer variability of various edge-angle measurement methods. The platform measurements for all 123 flakes for all five geometric methods are displayed in Figure 3. In these plots, the vertical dispersion of each set of six points represents the inter-observer variability and can be measured using the coefficient of determination \( R^2 \) values. Based on these coefficients of determination we can conclude that there is little variability in the dispersion of all five geometric approximations. With the poorest performing 2D approximation, the trapezoid, returning an \( R^2 \) value of 0.979, compared to 0.990 for the rectangle (a difference of 0.011), only a minimal increase in variability is observed. In fact, the triangle approximation values returned a slightly high \( R^2 \) value than the rectangular approximations (these results are summarised in Table 3.4).
Figure 3.5. Regressions of individual measurements conducted by six observers plotted against the mean of those measurements, with logarithmic axes. These charts examine the inter-observer variability of the GAP method among 6 participants.

To corroborate these results, we employ another method of inter-observer variability testing called the technical error of measurement (TEM), developed originally for anthropometric research (Ulijaszek and Kerr 1999; Weinberg et al. 2005), and introduced to archaeology by Lyman and VanPool (2009). The TEM can be calculated for the variability among more than two observers using the equation;
\[ TEM = \sqrt{\frac{\left(\sum_i^N \left(\sum_j^K M_j^2\right) - \left(\sum_j^K M_j\right)^2 / K\right)}{N(K-1)}} \]

where \( N \) is the sample size, \( K \) is the number of observers, and \( M \) is the measurement. The TEM is a measure of imprecision when multiple observers are involved. TEM values reveal the amount of variation that is caused by inter-observer variability rather than actual variability in the items being measured. Lower values denote less imprecision and values are expressed in the original measurement units.

**Table 3.4. Results of both measures of inter-observer variability.**

<table>
<thead>
<tr>
<th>Measure of Inter-User Variability</th>
<th>Rectangle</th>
<th>Triangle</th>
<th>Rhombus</th>
<th>Trapezoid</th>
<th>Ellipse</th>
<th>Average Values for Non-Rectangle Approximations</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R^2 )</td>
<td>0.990</td>
<td>0.992</td>
<td>0.984</td>
<td>0.979</td>
<td>0.982</td>
<td>0.984</td>
</tr>
<tr>
<td>TEM (mm(^2))</td>
<td>4.78</td>
<td>2.72</td>
<td>4.29</td>
<td>6.36</td>
<td>6.49</td>
<td>4.97</td>
</tr>
</tbody>
</table>

Table 3.4 shows the results of applying this equation to each of the five geometric approximations. Again, there is minimal difference among these TEM values. The non-rectangle approximations introduced as part of the GAP method return an average TEM value of 4.97mm\(^2\), compared with 4.78mm\(^2\) for the rectangle approximations. This is a difference of 0.19mm\(^2\), which is negligible when we consider that all 123 platforms had a mean platform area of 23.67mm\(^2\). In any case, the GAP method of platform measurement involves inter-observer variation between 2.72mm\(^2\) – 6.49mm\(^2\), compared with 4.78mm\(^2\) for the conventional method of platform width and thickness measurements. Based on these two measurements of inter-observer variation, there is no reason to reject the GAP method on the grounds of inter-observer variability, as barely any additional variability is involved in measuring the four new geometric approximations.

### 3.7 Discussion and Conclusions

The conventional method of platform area calculation, using platform width and thickness values, has been shown here to dramatically overestimate actual platform size. The longevity of this method can perhaps be attributed to the notion that while platform width and thickness measurements overestimate platform area, they do so to a consistent degree. It was perhaps presumed that rectangle approximation measurements would be inaccurate but precise. Some may argue that precise inaccuracy is more desirable than imprecise accuracy. This would be a valid argument in the interests of reproducibility, but this is not the pattern observed here. Instead, this study demonstrates wide ranging errors in the rectangular method, meaning that the status-quo of platform measurement is neither accurate nor precise.
As considered in the introduction, platform area is commonly and powerfully used to estimate reduction intensity, which in turn is used to develop complex models of human behaviour including technological organisation, land-use patterns, mobility, occupational intensity, raw-material availability and subsistence. With significant archaeological interpretations at stake, there is little room for a wholly inaccurate and imprecise method of platform measurement.

Like others (Braun et al. 2008a), we advocate here and elsewhere (Clarkson and Hiscock 2011; Muller and Clarkson 2014) that 3D scanning of flake attributes such as platforms is an ideal solution to minimise systematic measurement errors such as those involved in conventional platform measurements. Where time, resources or electricity is a constraining factor (as is so often the case in archaeology) the use of new geometric approximations of platforms offers significant increases in accuracy and precision.

A very recent study by Reti (2016) offers a comprehensive assessment of Oldowan morphological variation, in which platform area was an important variable. Tackling the same problem assessed here, the inadequacy of conventional platform measurements, Reti (2016) offers an alternative in the Supporting Information section. This approach involves measuring platform width and platform thickness at three different points each. These six measurements were used in conjunction with a coefficient and several equations to determine platform area. A regression of platform area taken from this new method against the values calculated from photographs of 40 flakes revealed a high coefficient of variation, providing an initial assessment of accuracy. However, regressions of measurements against known values are inadequate for differentiating the role of accuracy and precision in the efficacy of a new method, let alone teasing apart the influence of systematic versus random error (Lyman and VanPool 2009). Therefore, further research is required to test the accuracy, precision, inter-observer variability and efficiency of this promising method of platform area measurement.

With the accuracy and precision of the four new geometric approximations (triangle, rectangle, rhombus and ellipse) being statistically indistinguishable from the highly reliable 3D scanned measurements, we can be confident in the efficacy of the GAP method. While some subjectivity is introduced with the GAP method via the selection of which 2D shape best approximates the real platform, we demonstrated that even when no attempt is made to select the optimal 2D approximation, the GAP method still outperforms the conventional rectangular method in terms of both accuracy and precision.
The GAP method offers these significant improvements to precision and accuracy with close to nil increase in time invested in measurement and analysis. Furthermore, the method for converting existing platform width and thickness measurements of area to values compatible with measurements taken using the GAP method means that inter-site comparisons with previously analysed assemblages would not be hindered for those employing the GAP method. In addition, measurements from different sites become more translatable and reliable, and it becomes less likely for critical errors associated with interpretations relying on measurements from multiple sites. Finally, based on two measures of inter-observer variability there was no real increase in the variability associated with multiple users among any of the four new geometric approximations of platforms. Having presented and tested a new method of platform measurement, we find that the GAP method fulfils the need for an accurate, precise, flexible and quick measure of platform area where 3D scanning is unfeasible or impractical.
CHAPTER 4:

A new reduction intensity measure for backed blades: blank consumption, regularity and efficiency at the early Neolithic site of Boncuklu, Turkey


<table>
<thead>
<tr>
<th>Contributor</th>
<th>Statement of contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antoine Muller (Candidate)</td>
<td>Experimental design (90%)</td>
</tr>
<tr>
<td></td>
<td>Experimental sample (50%)</td>
</tr>
<tr>
<td></td>
<td>Analysis (70%)</td>
</tr>
<tr>
<td></td>
<td>Writing (90%)</td>
</tr>
<tr>
<td></td>
<td>Editing (40%)</td>
</tr>
<tr>
<td>Chris Clarkson</td>
<td>Experimental design (10%)</td>
</tr>
<tr>
<td></td>
<td>Experimental sample (50%)</td>
</tr>
<tr>
<td></td>
<td>Analysis (20%)</td>
</tr>
<tr>
<td></td>
<td>Writing (10%)</td>
</tr>
<tr>
<td></td>
<td>Editing (20%)</td>
</tr>
<tr>
<td>Douglas Baird</td>
<td>Archaeological sample (50%)</td>
</tr>
<tr>
<td></td>
<td>Analysis (10%)</td>
</tr>
<tr>
<td></td>
<td>Editing (20%)</td>
</tr>
<tr>
<td>Andrew Fairbairn</td>
<td>Archaeological sample (50%)</td>
</tr>
<tr>
<td></td>
<td>Editing (20%)</td>
</tr>
</tbody>
</table>

Abstract

Estimating the extent of reduction intensity on lithic artefacts has become a key goal of lithic analyses in recent decades. An understanding of a core or tool’s reduction intensity can lead to intra- and inter-site interpretations regarding patterns of human behaviour. Decades of research on this topic have led to a diverse suite of quantifiable measures of reduction encompassing almost all artefact types. We present and test the efficacy of a new method for measuring reduction intensity for backed blades, one of the few remaining artefact types lacking a reliable measure of reduction. As instances of successive retouching of backed blades are rarely documented, we use this
reduction intensity metric to simply estimate original blank size, rather than model multiple stages of reduction. Allometric relationships of blade thickness to length and width were used to estimate original microlith blank size, thereby offering a quantification of reduction intensity. This method was then applied to the microliths, a specific type of small backed blade, from the early Neolithic site of Boncuklu, Turkey. Reconstructing the original mass and dimensions of microlith blanks allows inferences to be made regarding the regularity, efficiency and decision making involved in microlith production. These microliths were produced using similarly sized blanks that were selectively and nearly completely consumed in this reduction sequence. This aspect of lithic technology at Boncuklu involved little waste and therefore enabled the inhabitants to cope with the constraints of raw material access and negotiate the changes to subsistence and social organisation occurring at the beginning of the Neolithic. While we test the efficacy of this new method on the Neolithic of Turkey, it could equally be applied to any blade- or bladelet-based backed artefact industry. These industries existed at various times in Europe, the Near East, Africa, South Asia, and parts of Australia, together spanning the Middle Stone Age, Later Stone Age/Upper Palaeolithic, Epipalaeolithic, Mesolithic and Neolithic.

4.1 Introduction

Since the concept of curation was introduced by Binford (1973; 1977a; 1979) it has received much attention in lithic research with a particular emphasis on developing methods of quantifying core and tool reduction intensity. This attention is due in part to the interpretive power of the concept. From an understanding of the extent of a core or tool’s curation, archaeologists can infer features of human behaviour, such as technological organisation, raw material consumption, mobility patterns and subsistence practices (Andrefsky 1994; 2009; Bamforth 1990; 1991; Binford 1973; 1977a; 1979; Blades 2003; Bleed 1986; Braun et al. 2008a; Close 1996; Dibble 1995; Hiscock and Attenbrow 2003; Odell 1996; Shott 2005; Shott and Ballenger 2007; Shott and Sillitoe 2004; 2005; Shott and Weedman 2007).

With reduction intensity forming an integral component of curation, there is an abundance of ‘universal’ and tool-specific reduction intensity metrics. Previous attempts at quantifying reduction intensity have relied on flake geometry (Eren et al. 2005; Eren and Sampson 2009; Hiscock and Clarkson 2005; 2009; Kuhn 1990; Morales et al. 2015), quantity and invasiveness of flake scars (Andrefsky 2006; Clarkson 2002; 2013; Clarkson et al. 2014), platform attributes (Braun et al. 2008a; Clarkson and Hiscock 2011; Davis and Shea 1998; Dibble 1995; 1997; 1998; Dibble and Pelcin 1995; Dibble and Rezek 2009; Dibble and Whittaker 1981; Dogandžić et al. 2015; Lin et al. 2013; Magnani et al. 2014; Muller and Clarkson 2014; 2016b; Pelcin 1997a; 1997b; 1997c; 1998;
Rezek et al. 2011; Shott et al. 2000), and flake allometry (Blades 2003; Goldstein 2014; Quinn et al. 2008). This ongoing pursuit of universal and tool-specific reduction measures has resulted in lithic analysts having a diverse suite of reduction intensity metrics at their disposal. Therefore, there are few core or tool types that are lacking an appropriate method of quantifying their level of reduction. Backed blades however, are one such tool type. Not only are there no specific measures of reduction intensity for backed blades, but none of the more ‘universal’ methods are suitable.

In its broadest sense, reduction intensity is a measure of the amount of stone removed from a core during knapping or from a tool during retouch. This can take the form of an absolute value of stone, or indices that quantify the relative amount of stone removed. Previous analyses of reduction intensity commonly focus on successive retouch events and the intensity of this retouch to reconstruct a tool’s use-life as an explanation for its morphology (Blades 2003; Dibble 1984; 1987; 1995; Hiscock and Attenbrow 2003; Hiscock and Clarkson 2007; Holdaway 1991; Holdaway et al. 1996; Kuhn 1995; Rolland and Dibble 1990; Shott 2005; Shott and Ballenger 2007). While the morphological variation of backed blades may at times be influenced by the extent of reduction (Barton and Neeley 1996; Neeley and Barton 1994), functional and/or cultural constraints are likely more significant (Bar-Yosef 1991; Fellner 1995; Goring-Morris 1996; Henry 1996; Kaufman 1995; Phillips 1996). Therefore, we are instead primarily interested in using a reduction intensity metric to reconstruct the original size of backed blades. The goal of this study is therefore not to reconstruct reduction intensity in its traditional sense, whereby a series of instances of retouching and reshaping are modelled. Rather we are interested in reduction intensity in its broadest sense, involving a reconstruction of how much stone is lost during reduction as well as the original size of the blank. Estimating the original blank size (mass, length and width) of backed blades can allow a reconstruction of the regularity and efficiency of blank consumption, as well as the technological choices involved in backed blade production. The efficacy of this method will be tested on the microliths from an early Neolithic site of Boncuklu, Turkey.

4.1.1 Towards a Measure of Reduction Intensity for Backed Blades and Bladelets

No existing reduction measure is suitable for backed blades or bladelets due to their specific morphology. Blades are defined as flakes which are twice as long as they are wide with parallel or slightly convergent margins and dorsal scars. Bladelets or microblades are defined following Tixier (1963), as blades whose maximum dimensions do not exceed 50mm in length and 12mm in width. Backed blades and backed bladelets are therefore any blade or bladelet respectively whose morphology has been modified via backing, or very steep retouch. Backed bladelets must not exceed 50mm in length or 9mm in width. As ‘bladelets’ are subsumed under the class of ‘blades’,
and in the interests of brevity, the terms ‘blades’ and ‘backed blades’ will be used in this paper as an inclusive term that encompasses their smaller variants, ‘bladelets’ and ‘backed bladelets’. Lastly, while a variety of definitions of microliths are at times conflated in the literature, for the purposes of this paper we define microliths (or geometric microliths) as specific versions of backed bladelets which typically form geometric shapes such as rectangles, trapezoids, triangles, crescents and intermediate variants. While this new method of measuring reduction intensity is tested on a sample of microliths from Boncuklu, the method is designed to encompass any backed blade or bladelet.

While the relationship between platform attributes and original flake size, originally employed by Dibble and Whittaker (1981), has proven a reliable measure of reduction intensity, backed blades very infrequently possess an extant platform. Methods based on flake scars (Andrefsky 2006; Clarkson 2002) are similarly unsuitable for backed blades, as backing typically occurs on one margin only, and due to the steep angle of backing, much variation in reduction intensity can exist while the number and invasiveness of flake scars varies modestly.

Geometric measures of reduction intensity, such as Kuhn’s (1990) Geometric Index of Unifacial Reduction (GIUR) and the Estimated Reduction Percentage (ERP) developed by Eren et al. (2005), potentially offer a more promising means of measuring reduction intensity on backed blades. Most estimates of reduction intensity require a quantification of the difference in artefact size (e.g. mass, volume or surface area) before and after retouch. In backed blade production, mass is typically removed from both the length and width dimensions, but seldom is there significant change in the medial thickness. Any reduction intensity metric therefore needs to be able to reconstruct the length and width lost during retouch.

As originally noticed by Dibble (1995), the GIUR is less sensitive to variability in loss of length for particularly long and flat flakes such as blades and bladelets. Kuhn’s (1990) GIUR is based on the ratio of retouch thickness (t) to flake thickness (T), ranging in values of 0 to 1, with 1 being the most heavily retouched. For elongate and flat flakes like blades, the maximum t/T ratio of 1 is typically achieved rapidly due to their minimal thickness, meaning that any further retouch will not result in further increases to the t/T ratio. While there has been disagreement regarding the magnitude of this phenomenon (Eren and Sampson 2009; Hiscock and Clarkson 2005; 2009), it poses serious limitations to the reliability of quantifying backed blade reduction intensity, as the manufacture process of these artefacts typically involves considerable loss of length. For these reasons, the ERP of Eren et al. (2005) is similarly unable to reliably reconstruct the amount of length lost during knapping.
While estimating lost length remains problematic for these geometric measures, the amount of width removed during retouch can potentially be estimated for lightly backed blades using the GIUR or ERP. When backing extends beyond all dorsal arrises however, neither the GIUR nor the ERP can be used to measure reduction intensity. An arris is defined as the intersecting ridge between two negative flake scars on the dorsal surface of a flake when this intersection occurs for the majority of the flake’s length. When backing extends beyond all arrises, the GIUR t/T ratio becomes 1 and the ERP ‘∠b’ value is unable to be calculated.

Unfortunately for those wishing to measure reduction intensity on backed blades, a large portion of such artefacts are heavily enough retouched to disqualify them from these geometric methods. To provide a gauge of how often this problem occurs, from the experimental sample of backed blades described below, purposely knapped to capture the widest range of reduction intensity possible, 37.9% of backed bladelets were backed invasively enough to remove all arrises. Additionally, of the backed bladelets from Boncuklu, 97.1% were too invasively backed to be suitable for such geometric measures.

While the width lost during retouch may be reconstructed for a portion of backed blades using geometric methods, the length remains unknown for all artefacts. As both the length and width lost during retouch needs to be quantified in order to estimate reduction intensity, we seek a more reliable measure for backed blades.

Having considered universal reduction intensity measures we now turn to allometric relationships of flake dimensions. Due to principles of fracture mechanics and limitations on the strength of stone that dictate flake allometry, for a blade to expand in one dimension it must expand in the other two dimensions. For example, a longer blade is typically wider and thicker than a shorter blade. These allometric principles have long been known (Dibble 1995; Holdaway 1991), with relationships between blade thickness and blade length, width and surface area serving lithicists seeking reduction intensity metrics. For example, Blades (2003) used the relationships of blade thickness to length ($r^2 = 0.462$) and surface area ($r^2 = 0.533$) to model hunter-gatherer mobility. More recently, Quinn et al. (2008) and Goldstein (2014) used the blade thickness and length relationship to estimate reduction intensity on el-Khiam points ($r^2 = 0.602$) and end-scrapers ($r^2 = 0.518$) respectively.
Although very promising, these approaches suffer from relatively low coefficient of determination values. This low predictive power of existing allometric models may be partly explained by low sample sizes, however normal variation in blade morphology is likely also to blame. A key task undertaken in this study is to find ways to better accommodate this variation in blade morphology, and therefore increase the predictive power of blade thickness allometry.

4.2 Experimental Method

In order to develop allometric relationships of thickness to length and width, a sample of 289 blades and bladelets were knapped by CC. These blades were produced with obsidian and flint, two of the most common materials for backed blade manufacture, especially in southwest Asia where Boncuklu is situated. Direct percussion, indirect percussion (punch) and pressure flaking were all used in order to examine the reliability of this method on multiple blade knapping techniques.

Ideally, these thickness to length and width regressions would be devised using unretouched blades from the site under investigation. However, due to the extremely high fragmentation rate at Boncuklu (94%), there were too few intact blades to form reliable regressions. Instead, an experimental knapper produced the replicative assemblage of 289 blade blanks based on the morphology of complete and incomplete blanks to provide a statistically robust sample. The range of blade making techniques observed at Boncuklu were used to ensure the experimental sample as closely as possible resembled the Boncuklu specimens. For sites with a sufficient sample of unretouched blade blanks, it is recommended that site-specific regressions of thickness to length and width be devised.

To better capture the relevant morphological variation of blades prior to backing, some crucial changes were made to the typical method of recording a singular measurement of flake width and thickness at the medial cross-section. This involved taking repeated measurements at regular intervals, ignoring the proximal and distal extremities that are almost invariably removed during backing, and incorporating additional variables such as edge angle and the number of arrises. This suite of digital calliper measurements included length, and three measurements taken at regular intervals of both medial width and medial thickness. Similarly, three edge angle values were recorded at regular intervals on both margins of the blade using a goniometer. The location of these three medial width, thickness and edge-angle measurements are represented in Figure 4.1. Raw material, termination type and the number of arrises were also recorded.
These measurements were then used to create allometric relationships of blade thickness to length and width. These allometric relationships were calculated using ordinary least square (OLS) regression. Although it is recommended to use reduced major axis (RMA) regression where the independent variable may contain measurement error and therefore violates the assumptions of OLS regression (Forstmeier 2011; Smith 2009; Sokal and Rohlf 2012; Warton et al. 2006), as is the case here, recent considerations of the two regression types found OLS to be more appropriate for allometric relationships where measurement error is manageably low (Al-Wathiqi and Rodriguez 2011; Kilmer and Rodriguez 2017). The utility of the relationships of blade thickness to length and width relies on the fact that the thickness of backed blades typically remains unchanged before and after backing. For example, in the experimental sample an average of only 5.67% of thickness was lost and the thickness of blades before and after backing were statistically indistinguishable ($t = 1.33$, d.f. = 263, $p = 0.19$). We therefore assume that the thickness of backed blades represents the thickness of the original blade blank. In archaeological scenarios where the original blank dimensions of backed blades are unknown, a backed blade’s thickness can be input into regressions for length versus thickness and width versus thickness to provide estimates of the percentage of length and width lost. This provides an estimate of original blank mass, which when compared to the actual backed blade mass, provides a quantifiable measure of backed blade reduction intensity.
To test the efficacy and reliability of this method, a sample of 132 of the original 289 blade blanks were randomly selected and retouched into microliths by AM. Microliths are a very common form of backed blade, and the most common tool type at Boncuklu. The measurements taken on the blanks were repeated for the microliths including the multiple measurements at regular intervals for width, thickness and edge angle. When possible, edge angle values were recorded for both lateral margins, but if backing was too invasive only the unretouched margin was measured. Similarly, the number of arrises was only recorded if they could be clearly seen or inferred from the cross sectional morphology. These measurements allowed the original blank mass to be predicted and compared to the actual microlith mass, providing a measure of reduction intensity. Summary data for both the blade blanks and the microliths are shown in Table 4.1.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Mass (g)</th>
<th>Length (mm)</th>
<th>Mean Medial Width (mm)</th>
<th>Mean Medial Thickness (mm)</th>
<th>Mean Medial Edge Angle (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>Mean</td>
<td>4.80</td>
<td>58.40</td>
<td>13.19</td>
<td>3.56</td>
</tr>
<tr>
<td>Blades</td>
<td>Std. Dev.</td>
<td>5.72</td>
<td>18.70</td>
<td>4.18</td>
<td>1.44</td>
</tr>
<tr>
<td>Experimental</td>
<td>Mean</td>
<td>2.01</td>
<td>34.62</td>
<td>10.09</td>
<td>3.40</td>
</tr>
<tr>
<td>Microliths</td>
<td>Std. Dev.</td>
<td>2.35</td>
<td>12.32</td>
<td>3.13</td>
<td>1.23</td>
</tr>
<tr>
<td>Boncuklu</td>
<td>Mean</td>
<td>0.19</td>
<td>15.41</td>
<td>4.89</td>
<td>1.57</td>
</tr>
<tr>
<td>Microliths</td>
<td>Std. Dev.</td>
<td>0.15</td>
<td>5.17</td>
<td>1.03</td>
<td>0.44</td>
</tr>
</tbody>
</table>

4.3 Experimental Results

As expected, our results confirm the general trend that blade length and width increase concomitantly with blade thickness. Table 4.2 shows these relationships and also demonstrates the positive effect on regression strength of the methodological changes outlined above. Using the typical singular width and thickness measurements, taken at the medial cross-section, the predictive power of thickness resembles previous attempts (Blades 2003; Goldstein 2014; Quinn et al. 2008) at using blade thickness to predict other blade dimensions ($R^2 = 0.657$ and 0.588 for length and width respectively). When the mean medial width and thickness values (average of three measurements taken at regular intervals) are used in the regressions however, the predictive power improves by approximately ten percentage points ($R^2 = 0.739$ and 0.686 for length and width respectively). While promising, these regressions suggest that only approximately 70% of variability in blade length and width is predicted by blade thickness alone.
Table 4.2. Regressions of the relationship of blade thickness to length and width, with associated equations and coefficient of determination \( (R^2) \) values. Values in bold represent the strongest regressions, and were therefore chosen to predict original length and width on the experimental and archaeological backed blades.

<table>
<thead>
<tr>
<th>Predicted Variable</th>
<th>X-Axis</th>
<th>Y-Axis</th>
<th>Equation</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>Medial Thickness</td>
<td>Length</td>
<td>( y = 9.82x + 22.12 )</td>
<td>0.657</td>
</tr>
<tr>
<td></td>
<td>Mean Medial Thickness</td>
<td>Length</td>
<td>( y = 11.13x + 18.74 )</td>
<td>0.739</td>
</tr>
<tr>
<td></td>
<td>Mean Medial Thickness /</td>
<td>Length</td>
<td>( y = 266x + 24.80 (1 \text{ arris}) )</td>
<td>0.557</td>
</tr>
<tr>
<td></td>
<td>Edge Angle</td>
<td></td>
<td>( y = 372x + 20.42 (2 \text{ arris}) )</td>
<td>0.579</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( y = 289x + 25.58 (\text{all}) )</td>
<td>0.515</td>
</tr>
<tr>
<td>Width</td>
<td>Medial Thickness</td>
<td>Medial Width</td>
<td>( y = 2.41x + 5.01 )</td>
<td>0.588</td>
</tr>
<tr>
<td></td>
<td>Mean Medial Thickness</td>
<td>Mean Medial Width</td>
<td>( y = 2.40x + 4.65 )</td>
<td>0.686</td>
</tr>
<tr>
<td></td>
<td>Mean Medial Thickness /</td>
<td>Mean Medial Width</td>
<td>( y = 77.66x + 3.14 (1 \text{ arris}) )</td>
<td>0.913</td>
</tr>
<tr>
<td></td>
<td>Edge Angle</td>
<td></td>
<td>( y = 99.24x + 3.31 (2+ \text{ arrises}) )</td>
<td>0.878</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( y = 80.02x + 4.10 (\text{all}) )</td>
<td>0.792</td>
</tr>
</tbody>
</table>

To further improve this predictive power, the variables of edge angle and cross-sectional morphology (number of arrises) were incorporated. The strongest regression involving thickness and width was achieved via plotting thickness against mean medial width divided by mean edge angle, and separating the blanks into those with one and two or more arrises. In cases where the number of arrises is apparent on the experimental and archaeological backed blades, the relevant equation can be used. When the number of arrises is unclear however, the equation that incorporates all data points can be used. Unfortunately, the length versus thickness regression could not be significantly strengthened using this multivariate approach, and thus the bivariate regression alone was used to predict original length. The strongest regressions, and therefore the equations that will be used to estimate original blade blank dimensions are charted in Figure 4.2.
Figure 4.2. Scatter plots showing the strongest regressions of the relationship of blade blank thickness to length (a) and width (b). The unbroken bold line in b represents values of all blade blanks and is the regression to be used when the number of arrises is uncertain.

Applying the appropriate regressions to each microlith allows an estimation of the percent of length and width removed during backing. Combining the percent of lost length and width can also provide an estimate of the percent of mass removed during backing and thus the original blank mass. On average, the 132 microliths were retouched by 50.54%, and the spread of this reduction intensity is shown in Figure 4.5 (left side). The roughly normal distribution is unsurprising, as the aim of the experimental blank production and backing was to produce the widest possible range of blank morphology and microlith reduction intensity, so as to test the versatility of this new method.

To visually explore the method’s reliability, the actual blank mass is plotted against this predicted blank mass in Figure 4.3 ($R^2 = 0.862$). As the data are positively skewed, a logarithmic scale was used on both axes. A two-sample paired t-test reveals a slightly significant difference between the actual and predicted values ($t = -2.37$, d.f. = 263, $p = 0.019$), likely representing the slight underestimation of this method. More importantly for gauging the reliability of a method, the percent error equation \( \frac{\text{actual} - \text{predicted}}{\text{actual}} \times 100 \) returns an average absolute error margin of 21.65%. With the accuracy of the prediction falling only marginally shy of the maximum accuracy (approximately 5-10%) for the best performing reduction intensity measures (see Hiscock and Tabrett 2010 for a comprehensive review of available metrics), this new metric appears relatively
reliable. This slightly lower accuracy is unsurprising, as flake allometry relies on trends of morphological variation within samples, whereas metrics relying on geometry or flake scar invasiveness are measuring tangible qualities of individual specimens. Having already discussed the incompatibility of backed blades with all other reduction intensity metrics, we propose that this new method could serve as a powerful means of estimating backed blade reduction intensity, if accompanied by an acknowledgement of the experimental error range. The microliths from Boncuklu provide an opportunity to test the archaeological utility of this new method.

![Figure 4.3. Scatter plot with logarithmic axes showing the actual original blank mass plotted against the predicted original blank mass.](image)

**4.4 Archaeological Case Study: Boncuklu**

Boncuklu (Baird et al. 2012; Baysal 2013a; 2013b; 2014; Fletcher et al. 2017; Spataro et al. in press) is an early Neolithic settlement mound in the Konya Plain of Central Anatolia, Turkey. The site is dated to 8,300-7,500 cal. B.C.E., contemporary with the later phases of occupation at the nearby Pınarbaşı (Baird et al. 2013; Baird et al. 2010; Fairbairn et al. 2014), and preceding the nearby Neolithic farming sites of Can Hasan III and Çatalhöyük (Ataman 1989; Baird 2012; Fairbairn 2005; Fairbairn et al. 2005; Hodder 2006). Subsistence practices at Boncuklu saw exploitation of birds, fish, tortoises and medium to large mammals, as well as a range of seeds, nuts and fruits. This predominantly forager lifestyle was supplemented by early instances of cultivation.
and herding (Baird et al. 2012). The lithic assemblage derives from a range of context types at the site, both mudbrick buildings and extensive midden deposits in open areas (Baird et al. 2012).

Consistent with its location in the middle of an alluvial environment with no immediately available stone sources, there is an extreme scarcity of stone raw material at Boncuklu, with all chipped stone artefacts analysed thus far amounting to less than 6kg. Geochemical sourcing has revealed that the obsidian exploited at Boncuklu originated from the lava flows of Göllü Dag and Nenezi Dağ in Cappadocia, Turkey, approximately 150km away. This distance was a likely constraint on raw material supply, with social factors relating to access to sources possibly creating further limitations. In any case, Boncuklu’s dearth of local raw material and position early in the Neolithic, at a time of significant changes to subsistence and social organisation, makes it an interesting case-study to examine reduction intensity and blank consumption patterns. Based on this archaeological context, it is hypothesised that the production of backed blades at Boncuklu likely contributed to preserving their limited supply of raw material, and that the modular functionality of backed blades assisted with the transition to small-scale food production.

The apparent limitation in access to lithic raw material at Boncuklu manifests in the core and tool strategies employed at the site via the efficient consumption of raw material. For example, the lithic artefacts at Boncuklu are remarkably small and possess very little cortex. The average lengths of complete flakes (N = 81, S.D. = 6.41, $\bar{x} = 11.13$mm), blades (N = 183, S.D. = 6.52, $\bar{x} = 15.96$mm) and cores (N = 31, S.D. = 7.89, $\bar{x} = 17.43$mm) are extremely small. Additionally, cortex or weathered surfaces are present on only 1.28% of artefacts, and only cover an average of 21.42% of the surface of those pieces. However, these measures only provide approximate estimates of lithic reduction intensity. To more reliably quantify these consumption patterns, we apply the method developed above on the microliths from Boncuklu. Backed blades, specifically microliths, are the most common formal tool type at Boncuklu making their reduction intensity integral to our analyses. Not only can this new method estimate the extent of reduction, it can also be used to reconstruct the original dimensions of microlith blanks from which patterns of blank consumption can be inferred.

Counts of blank, tool and core types are provided in Table 4.3. While the extreme small size of Boncuklu artefacts suggests that almost all early- and mid-stage knapping occurred off-site, the presence of cores, core trimming elements, spalls, microburins, and assorted small debitage suggests that later-stage knapping occurred at Boncuklu. With cores, blanks and tools being approximately equivalently small, it is likely that the majority of tools were manufactured on-site.
There are very few larger pieces that were likely knapped off-site, and these are predominantly interred in caches or alongside human burials. The lithic artefacts at Boncuklu are dominated by obsidian (96%), with flint and chalcedony infrequently present. Of the complete blanks analysed thus far from the Neolithic occupation of Boncuklu, blades (18%) and bladelets (51%) outnumber complete flakes. With a recent study finding bladelets to be one of the most efficient blank types (Muller and Clarkson 2016a), it is likely that the heavy reliance on bladelets contributed to the efficient use of raw material at Boncuklu. Of the formal retouched tools at Boncuklu, microliths ($N = 651$) are among the most common, but notches, burins, scrapers and piercers are also present. Microburins, a common by-product of microlith production, are present ($N = 202$) but likely were not involved in all instances of microlith production. A diverse range of microlith types exist at Boncuklu. Interestingly, this is starkly contrasted with the preceding Epipalaeolithic occupation at the nearby site of Pınarbaşı where all bar one microlith are lunates (Baird et al. 2013). Meanwhile, at the nearby and later site of Çatalhöyük, there are a few microliths predating 7,000 cal. B.C.E., but most chipped stone artefacts involve the manufacture and use of larger blanks (Carter 2007; 2011; Carter et al. 2008; Carter and Milić 2013; Conolly 1999a; 1999b).

**Table 4.3.** Counts of blanks, tools and cores analysed thus far at Boncuklu.

<table>
<thead>
<tr>
<th>Unretouched</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Flakes</td>
<td>8807</td>
</tr>
<tr>
<td>Blades/Bladelets</td>
<td>6688</td>
</tr>
<tr>
<td>Spalls</td>
<td>556</td>
</tr>
<tr>
<td>Core Trimming Elements</td>
<td>181</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Retouched</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pièces Esquillées</td>
<td>696</td>
</tr>
<tr>
<td>Microliths</td>
<td>651</td>
</tr>
<tr>
<td>Microburins</td>
<td>202</td>
</tr>
<tr>
<td>Backed or Truncated</td>
<td>469</td>
</tr>
<tr>
<td>Notches</td>
<td>172</td>
</tr>
<tr>
<td>Burins</td>
<td>167</td>
</tr>
<tr>
<td>Scrapers</td>
<td>80</td>
</tr>
<tr>
<td>Piercers</td>
<td>33</td>
</tr>
<tr>
<td>Points</td>
<td>28</td>
</tr>
<tr>
<td>Denticulates</td>
<td>9</td>
</tr>
<tr>
<td>Informal Tools</td>
<td>1552</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cores</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Platform</td>
<td>32</td>
</tr>
<tr>
<td>Opposed Platform</td>
<td>12</td>
</tr>
<tr>
<td>Multi-Platform</td>
<td>59</td>
</tr>
<tr>
<td>Bipolar</td>
<td>54</td>
</tr>
<tr>
<td>Core Fragments</td>
<td>121</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
</tr>
</tbody>
</table>

Two years of in-field lithic analysis yielded the pieces shown in Table 4.3, with a full suite of typological and technological classifications and measurements being conducted on a sample of
2,881 pieces. From this sample, 70 complete microliths have been identified, and these form the specimens that are the focus of this paper. Despite microliths being the most common formal tool type at Boncuklu, this relatively low number of complete specimens is explained by the extremely high fragmentation rate (94% for all pieces). These 70 microliths include scalene triangle (40%), triangle (17%), obliquely backed (13%), trapezoid (10%) and lunate (9%) pieces, as well as a range of non-geometric specimens (refer to Figure 4.4 for drawn examples and Table 4.1 for summary data).

**Figure 4.4.** Drawings of chipped stone from Boncuklu including microliths (a-f), microburins (g, h), scrapers (i-k) and a piercer (l). The solid line represents 1cm.
The Boncuklu microliths were analysed according to the methodology described above. These values were then entered into the allometric regressions, resulting in an estimate of the percentage of length and width lost from retouch, and thus the mass lost from retouch. The percentage of mass lost from retouch serves as the metric for reduction intensity. On average, 73.14% of original blade mass was lost during microlith production at Boncuklu, and the distribution of this reduction intensity can be seen in Figure 4.5 (right side). When compared to the experimental assemblage, which was knapped with the aim of providing a sample of maximum diversity, the Boncuklu microliths are retouched within a narrower range and with significantly higher intensity ($U = 1789.5$, d.f. = 201, $p < 0.001$).

A potential source of error in these results is the possibility of more than one microlith being produced from a single blank. As Goring-Morris (1996) points out, when calculating the feasibility of retrieving more than one microlith from a single blank, discarded blade blanks are a poor sample as they likely represent waste, especially in Boncuklu’s raw material poor environment. Herein lies another application of this method, as comparing microlith blank lengths (estimated using the above method) with microburin and microlith lengths provides a more reliable estimate of how many
microliths can be removed from a blank. For more than one microlith to be produced via the microburin method, the estimated blade blank must theoretically be longer than a microburin and two microliths. At Boncuklu, an average microburin ($\overline{x} = 10.01\text{mm}$) and two average microliths ($2\overline{x} = 30.62\text{mm}$) exceed the average length of the microlith blade blanks ($\overline{x} = 36.18\text{mm}$). However, a lower quartile microburin ($Q_1 = 8.01\text{mm}$) and two lower quartile microliths ($2Q_1 = 22.53\text{mm}$) do not exceed an average microlith blade blank ($\overline{x} = 36.18\text{mm}$), meaning that two small microliths could be theoretically retrieved from the microlith blanks.

In any case, the amount of microlith retouch is likely primarily determined by stylistic or functional constraints. For instance, microliths are often produced to be inserted into a haft, making the haft size a likely more significant determiner of reduction intensity. Additionally, while reduction intensity is often associated with economising behaviour, this link is difficult to make here. Not only are functional and stylistic constraints often involved in determining microlith reduction intensity, but the inherent small size of microliths means that the difference in the amount of stone wasted between a heavily retouched and a lightly retouched microlith amounts to little absolute mass. This is particularly apparent at Boncuklu where the microliths are especially small (Table 4.1). Therefore, for a more in depth application of this method, independent of the number of microliths removed from a blank, we now examine a model of blank consumption and microlith regularity at Boncuklu.

Figure 4.6 shows a bivariate boxplot (or ‘bagplot’) of the lengths and widths of complete Boncuklu blade blanks and microliths, as well as the estimated microlith blank size, which was calculated using the mass of Boncuklu microliths and the estimated percent of mass lost during retouching. The bivariate boxplot was first introduced by Rousseeuw et al. (1999) and is essentially a box-plot in two dimensions. The darker shaded areas are analogous to the ‘box’ of a box-plot, and the lighter shaded areas are analogous to the ‘whiskers’. Figure 4.6 was created with the ‘ggplot2’ package (Wickham 2009) using the R programming language (R Core Team 2015) and open access code written by Ben Marwick for bagplots (available at gist.github.com/benmarwick/00772ccea2dd0b0f1745).
Figure 4.6. Bivariate boxplot ('bagplot') showing the morphology of Boncuklu blade blanks and microliths in mm. The ‘estimated blade blanks’ category represents the length and width values obtained from the allometric regressions that converted Boncuklu microlith thickness values into estimates of the width and length of the microlith blanks prior to retouching.

Each aspect of this blank consumption model will be considered in turn, including the blank regularity, microlith regularity, efficiency and decision making involved in the utilisation of blade blanks at Boncuklu. Exploring the extent of blank and microlith regularity requires a measure of spread of these samples. While standard deviation and variance are useful measures of spread, they are reliant on samples possessing approximately equivalent means. The coefficient of variation (CV) on the other hand is independent of the magnitude of values as it is calculated by dividing the standard deviation by the sample's mean. For this reason, CV values have been used to quantify regularity and standardisation of lithic artefacts in scale-independent terms (Doelman and Holdaway 2011; Eerkens and Bettinger 2001; Low 2015; Mackay 2011). While several statistical methods for comparing CVs exist (Bennett 1976; Doornbos and Dijkstra 1983; Gupta and Ma 1996; Vangel 1996), the most reliable method for non-normal samples was put forth by Feltz and Miller (1996). This method has been applied archaeologically before (Eerkens 2000; Eren and Lycett 2016; Graf 2010; Lycett and Gowlett 2008; Okumura and Araujo 2014; Peelo 2011; Shipton et al. 2013b), and its use is advocated by Eerkens and Bettinger (2001) when comparing morphological standardisation or regularity of artefacts. Comparisons of CV values were conducted using the ‘cvequality’ package (Marwick and Krishnamoorthy 2016) in the R programming language (R Core Team 2015). We use the term regularity here to mean the extent of morphological similarity. Elsewhere, ‘standardisation’ is sometimes used to represent the same phenomenon, but bears
connotations with the enigmatic and difficult to quantify concept of intentionality. We therefore favour the term ‘regularity’ until a level of intentionality or predetermination can be demonstrated.

Figure 4.6 visually reveals that not all morphologies of blade blanks were used for microlith production at Boncuklu. When comparing the broad sample of blade blank (red) length and width values with the tighter clustering of estimated blade blank (green) values, only the very longest and narrowest blanks were selected. This visual comparison is confirmed when exploring the CV values for length of blade blanks and estimated blade blanks (53.79 versus 13.59 respectively), as well as their width values (47.00 versus 17.04 respectively). Feltz and Miller’s (1996) test for equality of coefficients of variation reveals that the spread of actual blade blanks is significantly larger than the blanks selected for microlith production (estimated blade blanks) in terms of both length (D’ AD = 68.31, d.f. = 215, p < 0.001) and width (D’ AD = 47.17, d.f. = 215, p < 0.001). This evidence suggests that the blanks selected for microlith production were highly regularised, with strict morphological constraints imposed on the selection of microlith blanks at Boncuklu.

Additionally, the microliths themselves appear to possess little morphological variation, with their length and width values clustering tightly. The microliths are also significantly smaller in terms of both length (U = 8, d.f. = 139, p < 0.001) and width (U = 344, d.f. = 139, p < 0.001) compared to the estimated original blade blank values. However, this approach overlooks the variation in the morphology of the blanks prior to retouch. Following the recommendations of Eerkens (Eerkens 1997; 1998; Eerkens and Bettinger 2001), we propose quantifying lithic regularity in scale-independent terms. In other words, tools can only be considered morphologically regular if they vary in size less than their blanks. Otherwise, exploring only tool morphological variation could potentially misidentify the source of regularity, or the stage at which regularity was applied in the reduction sequence. For instance, were strict morphological constraints imposed on the blanks, or on the tools, or both? To examine whether morphologically regular blanks were selected, or whether the microliths were morphologically regulated during retouch, the coefficient of variation (CV) values for length and width can be compared.

While the morphology of both the estimated blade blanks and microliths vary minimally, CV values of the estimated blade blanks are smaller than the microliths for both length (13.59 and 33.77 respectively) and width (17.04 and 23.35 respectively). Feltz and Miller’s (1996) test for equality of coefficients of variation reveals that the scale-independent spread of microlith size is significantly greater than the estimated blade blanks, in terms of both length (D’ AD = 45.09, d.f. = 139, p < 0.001) and width (D’ AD = 6.23, d.f. = 139, p = 0.01). While the microliths appear to be regular,
varying minimally in length and width, further morphological regularity does not appear to be applied to the microliths beyond the strict morphological constraints imposed on blank selection. Therefore, in scale-independent terms the microliths at Boncuklu bear little morphological regularity. The narrow range of morphology appears to be imposed during blank selection, rather than microlith backing.

Examining the relative distribution of blade blank and estimated blade blank values also informs the efficiency of the consumption of blanks at Boncuklu, or how completely blanks of suitable morphology were selected for microlith production. There appears to be an almost complete consumption of blanks within the range of approximately 5-10mm of width and 30-50mm of length. As discussed above, this range appears to be the morphology of blanks selected for microlith production by the inhabitants of Boncuklu. Only five blade blanks overlap in morphology with the blanks chosen for microlith production (estimated blade blanks), and none overlap in the darker shaded area (representing the central 50% of the sample). Statistically, the estimated blade blanks are significantly longer ($U = 637$, d.f. = 212, $p < 0.001$) and narrower ($U = 2195$, d.f. = 212, $p < 0.001$). These results suggest a very efficient consumption of raw material, with almost all blanks of suitable size and shape being utilised in the microlith knapping schema.

Collectively, this evidence can be used to infer possible instances of intentionality in the microlith blank consumption at Boncuklu. It appears that the inhabitants of Boncuklu precisely selected similarly sized blanks for microlith production while ignoring blanks outside this desired range. However, were they also concertedly attempting to produce similarly sized blanks reserved for microlith production? The extremely high proportion of microburins possessing platform preparation provides evidence that both selection and predetermination occurred. Platform preparation occurs on a core’s platform and flaking surface and can take the form of delivered strikes (overhang removal on the core surface and faceting on the platform) or grinding. As mentioned earlier, microburins are involved in microlith production and are typically discarded without any further use. Of the 39 complete microburins at Boncuklu, all are proximal and only six have crushed platforms. From the remaining 33 microburins, all possess evidence of either overhang removal or faceting and 79% possess grinding. When compared to the blade and bladelet blanks in the assemblage that are either complete or retain their proximal end, only 79% possess overhang removal or faceting and 41% possess grinding (Figure 4.7). The different proportions of platform preparation and faceting ($\chi^2 = 8.41$, d.f. = 3, $p = 0.038$), and grinding ($\chi^2 = 14.37$, d.f. = 3, $p = 0.002$) are statistically significant between the microburins and blade blanks.
Platform preparation is aimed at removing unwanted stone from cores and contributes to successful core reduction. Unprepared cores are far more likely to result in minor and major flaking flaws that inhibit further blank production. Importantly also, a high level of blank morphological predetermination necessitates high levels of platform preparation. For these reasons, the presence of microburin platform preparation at much higher rates than blade blanks suggests a level of predetermination of microlith blank removals. It appears that the Boncuklu knappers attempted to produce blanks of the morphology shown in green in Figure 4.6, and that blanks outside of this narrow range were overlooked for microlith production. This evidence can be contrasted with cases where flakes were likely selected post-knapping based on desired morphology as part of an expedient core and flake technological schema (e.g. Borel et al. 2013; Flenniken and White 1985). Combined with the microlith blank regularity and consumption explored above, the manufacture of microliths appears to be a concerted and predetermined strategy involving a narrow window of the reduction sequence.

Figure 4.7. Stacked bar charts showing the presence and location of platform preparation, including overhang removal and faceting (a), and grinding (b) for microburins and blade blanks with intact platforms.
4.5 Discussion and Conclusions

This paper developed and tested a new method for estimating reduction intensity for backed blades, a tool type that previously had no reliable reduction intensity metric available. Based on the experimental and archaeological applications of this method considered here, the accuracy of this method appears to be confirmed, with levels of accuracy approaching the best available measures of reduction intensity. The case study of Boncuklu in Turkey provides a verification of the potential efficacy of this method, but once applied to other backed blade assemblages the full utility and limits of this method can be more comprehensively realised. This new method could also be suitable for any other blade-based technology currently lacking a reliable measure of reduction. While most backed artefacts are likely retouched or backed only once, this method could be applied to tool types that are typically modified over successive stages of retouch, such as scrapers and a range of point types. Applying this method to successively retouched items would allow a reconstruction of the reduction intensity during sequences of retouch and re-shaping; a more traditional use of reduction intensity metrics. The application of the method in this present paper however, reconstructs reduction intensity in its broadest and simplest sense. Namely, it reconstructs how much stone was lost during backed blade production, thereby estimating the original mass and dimensions of the blade blanks.

While this new method for estimating the amount of stone lost during backed blade production has been tested on the microliths from an Anatolian Neolithic assemblage, it could equally be applied to any backed blade or bladelet based assemblage, such as the other various facies of the Near Eastern Epipalaeolithic and Neolithic (Albrecht 1988; Baird et al. 2013; Bar-Yosef 1998; Belfer-Cohen and Goring-Morris 2002; Carter 2011; Carter and Milić 2013; Grosman 2003; McDonald 1991; Neeley 2002; Neeley and Barton 1994; Todd 1966). The European Upper Palaeolithic and Mesolithic also provide an assortment of blade based backed artefacts (Bachechi et al. 1997; Kuhn 2002; Straus 2002), as does the Late Pleistocene and Holocene sequences of India and Sri Lanka (Abeyratne 1994; Clarkson et al. 2009; James and Petraglia 2005; Petraglia et al. 2009a; Petraglia et al. 2009b; Roberts et al. 2015). Finally, the microliths and backed bladelets of the African LSA and Epipalaeolithic (Ambrose 2002; Barton et al. 2013; Bouzouggar et al. 2008; Olszewski et al. 2011; Wadley 1993), as well as those microliths of the South African Howiesons Poort that were made on blade blanks (Clarkson 2010; Lewis et al. 2014; Soriano et al. 2015; Soriano et al. 2007; Villa et al. 2010; Wurz 1999; Wurz and Lombard 2007), offer an opportunity to apply this method. More broadly, this method could serve any blade based tool-type currently lacking a reliable reduction intensity metric. For example, while end-scrapers often possess an extant platform, which can be
used to model reduction intensity, double-end scrapers have remained a troublesome tool-type for estimating reduction intensity and could be served using this new method.

Applying this new method to the microliths of Boncuklu provides a model of microlith blank consumption. Reconstructing the original mass and dimensions of microlith blanks allows inferences to be made regarding microlith blank regularity, microlith regularity and the efficiency of blank consumption. These analyses culminate in an appreciation of the possible decision making involved in blank production and microlith backing occurring at Boncuklu. This blank consumption model, coupled with the platform maintenance of microburins, suggests that more investment was devoted to knapping the long and narrow blanks that are used in the microlith reduction sequence. This raises the likelihood that some blade blanks were made with the express purpose of later backing to form microliths. Additionally, increased platform preparation may increase the likelihood of removing sufficiently long and narrow blanks, with the applied force being more efficiently distributed into the core. In any case, there was a seemingly intentional and selective choice of these regularised blanks for microlith manufacture. This potentially intentional morphological similarity of microlith blanks raises the possibility that these blanks were not only regularised, but also standardised.

Interestingly, the microliths themselves possessed low regularity relative to the regularity of the microlith blanks. Despite little investment in microlith regularity, the high reduction intensity (an average of 73.14% of mass was removed during retouch) suggests that much investment was devoted to the final form of microliths, creating specific, but not regular morphologies. Not only were microliths morphologically diverse, they were also typologically diverse, with scalene, triangle, obliquely truncated, trapezoidal and lunate microliths all present. Comparatively, the microliths from the Epipalaeolithic phases of the nearby site of Pınarbaşı bear much greater typological regularity, with lunates comprising almost all microliths (Baird et al. 2013). This greater microlith morphological and typological variability at Boncuklu may reflect a higher number of functions being served by the Boncuklu microliths compared with those from Pınarbaşı. The floral and faunal evidence at these sites appears to substantiate this hypothesis. The Epipalaeolithic phases of Pınarbaşı were dominated by hunter-gatherer practices, while at Neolithic Boncuklu, small-scale food production such as cultivation and herding began to supplement the continuing hunter-gatherer practices (Baird et al. 2012). The increasing diversity of subsistence practices therefore approximately aligns with the diversification of microlith technology. The modular nature of microliths makes them an inherently flexible tool type, useful in a situation where the range and type of tasks carried out by the community were undergoing significant change, adjusting to the
inclusion of small-scale food production in subsistence practices. Their presence in scenarios of both foraging and small-scale cultivation and herding further attests to their flexibility.

Aside from their flexibility, microliths also contributed to the efficiency of the Boncuklu tool-kit within the context of limited access to raw materials at the site. The efficiency of microlith production provides one regionally long-standing strategy that served to conserve the limited quantity of raw material. While reduction metrics typically inform technological efficiency by reconstructing how much stone was lost during knapping, this is likely irrelevant here. As microliths are intrinsically very small, the amount of stone lost from a heavily retouched microlith compared with a lightly retouched microlith represents little actual mass difference. Despite an average of 73.14% of original blade mass being removed during microlith production at Boncuklu, the average mass of complete microliths is 0.20g, meaning that lighter retouch would confer almost nil economic advantage. In microlith production, far greater gains in efficiency can be attained by wasting as few blanks as possible.

The blank consumption model considered above revealed that very few blanks of suitably long and narrow morphology escaped the microlith reduction schema, making microlith production an efficient component of the technological organisation at the site. The near complete utilisation of suitable blanks is not the only aspect of the microlith knapping schema that contributed to raw material conservation, with microliths being identified as inherently efficient. For example, they are viewed as durable, reliable and maintainable components of an efficient and portable toolkit that offset the risk from environmentally, demographically, economically or socially driven stress (Clarkson et al. in press; Hiscock 1994; 2002; Hiscock et al. 2011; Neeley 2002). Microliths are also made on very small blades, or bladelets. Recent evidence suggests that blade knapping in general, and bladelets in particular, are the most efficient knapping strategies and products in terms of sharp edge length per gram of core (Muller and Clarkson 2016a). Bladelet and microlith production boasts a long-standing history in the broader region, dominating much of the earlier Epipalaeolithic assemblages. This abundance of microliths may therefore be predicated on cultural norms and traditions (c.f. Barton and Neeley 1996; Neeley and Barton 1994). Whether microliths were produced for functional or cultural reasons however, their innate flexibility and efficiency make them a functionally beneficial addition to any tool-kit. It is possible for efficiency to be an unknown aspect of a technology, whereby an adaptive economic advantage is conferred without efficiency being the primary conscious determiner of typology or morphology. With blades and bladelets comprising approximately 69% of complete blanks at Boncuklu, intentionally or not this high prevalence of blade technology contributed to the efficiency of knapping at the site.
This efficient production of microliths on blades, an already efficient blank type, is likely one of many strategies employed at Boncuklu to enable existence in a scenario of limited access to lithic raw materials. Meanwhile, the versatility and diversity of microliths likely facilitated the continuation of hunting and gathering practices while also meeting the diverse functional requirements demanded by the introduction of small-scale cultivation and herding. In summary, the newly introduced backed blade reduction intensity and blank consumption metric has revealed that this microlith knapping schema constitutes one technological strategy that enabled the inhabitants of Boncuklu to navigate the dynamic transitions occurring at the beginning of the Neolithic.
CHAPTER 5:
Identifying Major Transitions in the Evolution of Lithic Cutting Edge Production Rates


Abstract
The notion that the evolution of core reduction strategies involved increasing efficiency in cutting edge production is prevalent in narratives of hominin technological evolution. Yet a number of studies comparing two different knapping technologies have found no significant differences in edge production. Using digital analysis methods we present an investigation of raw material efficiency in eight core technologies broadly representative of the long-term evolution of lithic technology. These are bipolar, multiplatform, discoidal, biface, Levallois, prismatic blade, punch blade and pressure blade production. Raw material efficiency is assessed by the ratio of cutting edge length to original core mass. We also examine which flake attributes contribute to maximising raw material efficiency, as well as compare the difference between expert and intermediate knappers in terms of cutting edge produced per gram of core. We identify a gradual increase in raw material efficiency over the broad sweep of lithic technological evolution. The results indicate that the most significant transition in lithic production efficiency likely took place with the introduction of small foliate biface, Levallois and prismatic blade knapping, all introduced in the Middle Stone Age / Middle Palaeolithic among early *Homo sapiens* and Neanderthals. This suggests that no difference in technological efficiency existed between these species. With prismatic blade technology securely dated to the Middle Palaeolithic, by including the more recent punch and pressure blade technology our results dispel the notion that the transition to the Upper Palaeolithic was accompanied by an increase in technological efficiency. However, further increases in cutting edge efficiency are evident, with pressure blades possessing the highest efficiency in this study, indicating that late/epi-Palaeolithic and Neolithic blade technologies further increased efficiency.

5.1 Introduction
Technological efficiency is a key aspect of palaeoanthropological debates surrounding such topics as cognition, skill, intentionality, modernity, technological organisation and technological diversity (Brantingham and Kuhn 2001; Brown et al. 2009; Brown et al. 2012; de la Torre 2004; de la Torre et al. 2003; Ludwig and Harris 1998; Machin et al. 2007; Nonaka et al. 2010; Stout 2011; Stout et
al. 2014; Toth et al. 2006). It is commonly argued that innovations in lithic technology over the sweep of human evolution were accompanied by greater striking precision, longer reduction sequences, finer retouch, greater recursion and hierarchical planning, a greater variety of percussive and pressure flaking techniques, more intensive platform preparation, and predetermined and more standardised end-products (Cole 2015; Goren-Inbar 2011; Gallowt 1984; 1988; 2011; Heighway 2011; Moore 2010; Pelegrin 2009; Shipton et al. 2013b; Shipton 2013; Stout 2011; Stout et al. 2014; White et al. 2011; Wynn 2002). These technological changes are also often viewed as existing in a feedback loop with bio-morphological evolution that drove dexterity, cognition, and syntactic language (Greenfield 1991; Higuchi et al. 2009; Mahaney 2014; Morgan et al. 2015; Stout 2011; Stout and Chaminade 2007; 2009; 2012; Stout et al. 2011; Stout et al. 2008; Uomini and Meyer 2013; Wynn 2002). Within this narrative, blade and microblade technologies are often depicted as the pinnacle of evolution in core technology and a key component of the ‘Upper Palaeolithic Revolution’, involving highly standardised blank production and careful preparation and maintenance of core volume and efficiency (Ambrose 2001; Bordaz 1970; Brown et al. 2012; Klein 1989; Mellars 1989a; 1989b; Price 2007; Sheets and Muto 1972; Sherratt 1997).

Of particular concern to this study is this pervasive assumption that blades offer greater efficiency in cutting edge production (Bordaz 1970; Collins 1999; Klein 1989; Leroi-Gourhan 1957; 1993; Marks and Chabai 2006; Renfrew and Bahn 2012; Schick and Toth 1993; Whittaker 1994), underpinned by an early experiment examining the efficiency in edge production of pressure blade cores (Sheets and Muto 1972). Those who are not convinced of the gains in efficiency offered by blade production cite the raw-material wastage involved in selecting high-quality stone required for successful blade manufacture, the higher risk of critical breakages owing to the thinness of blades, and the fewer opportunities for retouch events due to the narrowness of blades (Bar-Yosef and Kuhn 1999; Chazan 1995; Hayden et al. 1996).

Stone knapping technologies are often portrayed as evolving in a linear fashion, described by Clark (1969) as a series of ‘modes’. The sequence begins with the single and multiplatform cobble industries of the Oldowan (Mode 1) at c.2.6 million years ago (mya), developing into bifacial and discoidal technologies (Mode 2) of the Early Stone Age/Lower Palaeolithic after c.1.6 mya. These were followed by the first appearance of Levallois (Mode 3) in the Middle Stone Age/Middle Palaeolithic, the development of blade technology (Mode 4) in the Upper Palaeolithic, and finally the appearance of the microlithic industries (Mode 5) of the Later Stone Age and Mesolithic. Despite the popularity of this scheme, it is now clear that technological evolution is far from linear, but is instead multidirectional, branching and recursive. For example, blade technology is securely
dated to well before the Upper Palaeolithic (Deino and McBrearty 2002; Johnson and McBrearty 2010; Shimelmitz et al. 2011; Soriano et al. 2007; Wilkins 2012), is not confined to anatomically modern humans, and appears and disappears in many regions over time (Bar-Yosef and Kuhn 1999; Beck and Jones 2015; McBrearty and Brooks 2000; Wendorf and Schild 1974).

Several experiments over the last four decades have compared raw material efficiency for a range of core reduction strategies (Chazan 1995; Eren et al. 2008; Heighway 2011; Jennings et al. 2010; Prasciunas 2007; Rasic and Andrefksy 2001; Sheets and Muto 1972), most of which consider the efficiency of blade core reduction. All bar one (Sheets and Muto 1972) of these experiments have called into question the supposed advantages in efficiency afforded by blade technology. Despite these findings, these experiments typically involve a comparison of only two reduction strategies such as biface versus blade or discoidal versus blade for example. For this reason, this paper compares the efficiency of eight core reduction strategies (bipolar, multiplatform, discoidal, biface, Levallois, prismatic blade, punch blade and pressure blade), which are common throughout the span of human evolution. We therefore provide the most comprehensive study of core efficiency to date. Previous experiments have also employed varied methodologies, hampering direct comparison of results. We therefore adopt the method of computer analysis developed by Eren et al. (2008) to measure cutting edge to mass ratios.

5.1.1 Core Reduction Efficiency

Sheets and Muto (1972) initiated research into core reduction efficiency by demonstrating the efficiency of pressure blades in terms of cutting edge per gram of core. Their method for calculating the cutting edge length, by measuring the length of the blade and doubling the result, was inaccurate considering that length measurements do not account for wavy or tapering blade edges, and their assumption of blade symmetry introduces a high degree of error.

More recently, some have sought to experimentally assess the raw material efficiency of biface reduction (Jennings et al. 2010; Prasciunas 2007; Rasic and Andrefksy 2001). Rasic and Andrefksy (2001) and Jennings et al. (2010) compared blade cores to bifacial reduction, finding parity in their raw material efficiency. These analyses did not include a consideration of cutting edge length however, focussing instead on blank count, size and shape. Cutting edge length was considered in a study by Prasciunas (2007), who found bifacial and multiplatform reduction to be equivalently efficient when considering blanks only larger than 5g. While each of these experiments highlight the efficiency of bifaces, and bring into question the supposed advantages in efficiency afforded by blade reduction, the variety of methods and units of measurement used to assess raw material
efficiency hamper comparisons among these experiments, and between the earlier work of Sheets and Muto (1972). Additionally, the use of a range of percussor types, such as soft and hard hammers, or hammers of different sizes, limits the reproducibility of these studies as different percussors can influence core and flake morphology (Crabtree 1968).

Brantingham and Kuhn (2001) applied geometric models to Levallois core reduction and found that the nature of Levallois reduction is geared towards minimizing waste and maximizing productive output. We therefore include Levallois flaking in our experimental sample to test this hypothesis in relation to the other technologies and situate it in the broad sweep of technological evolution.

Another methodologically rigorous approach to raw material efficiency was conducted by Eren et al. (2008), who compared the cutting edge length per original core mass of prismatic blade technology against discoidal technology, finding no significant difference between blade and discoidal cores. The hypothesis of Chazan (1995) that wider flakes can more frequently be resharpened thereby extending their use-life was also tested by Eren et al. (2008), who found that when the potential for further retouch events is considered, discoidal reduction is more efficient in terms of cutting edge per gram of core than blade reduction. Also of interest to this present study is the highly precise and reproducible method of Eren et al. (2008), who measured cutting edge length by reducing photographs to complex polygons and employing software to calculate the edge length. In the interests of reproducibility and comparability of results, this approach is also adopted here.

While Eren et al. (2008) set out to examine the transition to the Upper Palaeolithic using prismatic blade technology, more recent dates situate the advent of this technology well before the Upper Palaeolithic (Deino and McBrearty 2002; Johnson and McBrearty 2010; Shimelmitz et al. 2011; Soriano et al. 2007; Wilkins 2012). Our sample includes punch and pressure blade technology, which hitherto have only been dated to the Upper Palaeolithic and onwards. Thus, while Eren et al. (2008) were in effect comparing the Lower to Middle Palaeolithic transition, we offer the first real examination of cutting edge efficiency beyond the Middle Palaeolithic and into the Upper Palaeolithic, Epipalaeolithic and Neolithic (also including late Mesoamerican technologies). Moreover, while all previous attempts at examining core efficiency compared no more than two technologies, we compare eight technologies that broadly represent the evolution of core technology from the Oldowan to the Neolithic. As these previous studies found raw material efficiency to be equivalent among bifaces and prismatic blade cores (Jennings et al. 2010; Rasic and Andrefksy 2001), bifaces and multiplatform cores (Prasciunas 2007), and prismatic blades and
discoidal cores (Eren et al. 2008), we seek to test the null hypothesis that no significant differences in cutting edge per gram occur among the eight different reduction strategies under investigation.

5.1.2 Causes of Variability in Efficiency
A previous attempt at identifying the features of a flake which maximise its usable edge per unit of volume was conducted by Lin et al. (2013), who found that increasing the size of flakes, increasing the ratio of length to width (elongation), decreasing flake thickness relative to surface area, and decreasing platform size, could all contribute to maximising the efficiency of individual flakes. Moreover, they argue that these features can be maximised for flakes by decreasing platform depth and increasing exterior platform angle (EPA). The large sample produced in this present study allows for a consideration of the role of these, and other, flake attributes in altering cutting edge efficiency and the tendency of different common and well-known reduction sequences to increase cutting edge efficiency by emphasising these features.

5.1.3 Skill
We additionally examine the effect of knapping skill on the efficiency of reduction sequences, a divergence from previous knapping skill studies which typically focus on core reduction ability. Previous approaches to knapping skill include considerations of the presence of successes or failures in the knapping sequence (Bleed 2008; de la Torre 2004; Delagnes and Roche 2005), experimental attempts to identify markers of knapping skill in the individual (Eren et al. 2011b), ethnographic reconstructions of complex knapping sequences (Stout 2002), and analyses of the effect of raw material quality on knapping skill (Brantingham et al. 2000; Eren et al. 2011c). In experiments and the archaeological record, successive step or hinge terminations, overshot flakes, flakes with an undesired morphology, percussor marks attempted too far from the platform edge or on platforms of unsuitable angles have all been used as evidence of comparatively unskilful knapping (de la Torre 2004; Delagnes and Roche 2005; Eren et al. 2011b; Eren et al. 2011c; Finlay 2008; Geribâs et al. 2010; Harmand et al. 2015; Nonaka et al. 2010; Shelley 1990). Of particular interest to this study is that cores knapped by novice or intermediate knappers tend to have a higher rate of unsuccessful flake removals and produce flakes of smaller size (Finlay 2008; Geribâs et al. 2010; Nonaka et al. 2010; Shelley 1990; Toth et al. 2006). The influence of this discrepancy between intermediate knappers and experts on efficiency of cutting edge length per gram of core will be explored in this study.
5.2 Materials and Methods

5.2.1 Knapping Experiments

A total of 44 cores were knapped in this experiment to determine the efficiency of each reduction strategy. Reduction efficiency was defined as the length of resulting cutting edge relative to original core mass. As Eren et al. (2008) already examined the role of use-life on raw material efficiency, we consider efficiency in terms of cutting edge length per gram of original core for unretouched flakes only. The sample of technologies examined in this study includes bipolar, multiplatform, discoidal, bifacial, Levallois, prismatic blade, punch blade and pressure blade technology. As we also seek to evaluate the role of knapping skill on cutting edge efficiency, both expert and intermediate knappers were involved in most of these reduction sequences. The expert knapper has approximately two decades of experience in stone knapping, while the intermediate knapper has only a few years of experience but could adequately reproduce technologies like Levallois and prismatic blade reduction. The expert knapper reduced two cores and the intermediate knapper reduced five cores for the multiplatform, discoidal, biface, Levallois and prismatic blade technologies. The intermediate knapper did not conduct the bipolar portion of the experiment as this technology requires such little skill that minimal variation in cutting edge efficiency is expected. Similarly, punch and pressure blade knapping requires such a high level of skill that it could be executed only by the expert knapper. For each of these three technologies, the expert knapper conducted three repetitions.

The results produced from both the intermediate and expert knapper are included in all analyses of cutting edge efficiency in an attempt to capture the broad spectrum of skill among past hominin knappers. Including only an expert *Homo sapiens* knapper would not adequately summarise the millennia of evolution in brain (Frey 2008; Greenfield 1991; Stout and Chaminade 2012; Wynn 2002) and hand (Marzke 1997; 2013; Rolian et al. 2011) morphology that influences the cognition and skill of different knappers in the past and present.

All nodules used in this experiment were of the same high-quality flint with a high proportion of remnant cortex, weighing approximately 700g (Table 5.1), with the exception of the three bipolar cores that were far smaller owing to the fact that bipolar reduction typically occurs only on small cores. A Kruskal-Wallis test reveals that there is no significant difference between the mean of core masses for any reduction strategy other than bipolar (H = 3.18, d.f. = 6, 34, p = 0.78). All flakes were detached using the same standardised copper-headed billet weighing roughly 140g. This modern billet was used in favour of more traditional billets as the mass and hardness of copper is analogous to soft stone, antler or wood (Clark 2012; Crabtree 1967; 1968; Sheets and Muto 1972),
and the copper billet provided a constant and standardised shape throughout all twenty-eight experiments. All knapping debris was collected for later analysis.

**Table 5.1.** Mass values of the bipolar, multiplatform, discoidal, biface, Levallois, prismatic blade, punch blade and pressure blade cores, waste and blanks from each reduction sequence. Initial nodule masses in bold refer to reduction sequences conducted by the expert knapper.

<table>
<thead>
<tr>
<th>Core</th>
<th>Initial Nodule Mass (g)</th>
<th>Exhausted Core Mass (g)</th>
<th>Waste Chips Mass (g)</th>
<th>Total Blanks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bipolar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>119</td>
<td>4.54</td>
<td>12.41</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>134</td>
<td>4.08</td>
<td>16.13</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>152</td>
<td>35.90</td>
<td>29.01</td>
<td>28</td>
</tr>
<tr>
<td>Multiplatform</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>722</td>
<td>13.5</td>
<td>50.8</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>766</td>
<td>17.0</td>
<td>43.7</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>720</td>
<td>67.9</td>
<td>145.7</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>704</td>
<td>37.1</td>
<td>139.3</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>690</td>
<td>12.4</td>
<td>68.1</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>740</td>
<td>23.2</td>
<td>318.5</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>725</td>
<td>24.9</td>
<td>148.1</td>
<td>143</td>
</tr>
<tr>
<td>Discoidal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>790</td>
<td>17.5</td>
<td>89.5</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>772</td>
<td>9.0</td>
<td>75.2</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>741</td>
<td>46.3</td>
<td>104.6</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>709</td>
<td>39.7</td>
<td>167.1</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>690</td>
<td>44.7</td>
<td>140.2</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>730</td>
<td>137.8</td>
<td>235.8</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>712</td>
<td>55.0</td>
<td>165.4</td>
<td>145</td>
</tr>
<tr>
<td>Biface</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>747</td>
<td>32.42</td>
<td>187.53</td>
<td>226</td>
</tr>
<tr>
<td></td>
<td>786</td>
<td>34.77</td>
<td>150.7</td>
<td>220</td>
</tr>
<tr>
<td></td>
<td>757</td>
<td>13.19</td>
<td>152.96</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td>703</td>
<td>20.92</td>
<td>124.32</td>
<td>198</td>
</tr>
<tr>
<td></td>
<td>716</td>
<td>7.83</td>
<td>99.73</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td>754</td>
<td>29.15</td>
<td>120.88</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td>677</td>
<td>28.01</td>
<td>103.96</td>
<td>194</td>
</tr>
<tr>
<td>Levallois</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>758</td>
<td>10.8</td>
<td>142.7</td>
<td>134</td>
</tr>
<tr>
<td></td>
<td>724</td>
<td>15.1</td>
<td>122.5</td>
<td>147</td>
</tr>
<tr>
<td></td>
<td>712</td>
<td>76.7</td>
<td>176.9</td>
<td>167</td>
</tr>
<tr>
<td></td>
<td>740</td>
<td>48.7</td>
<td>171.8</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>704</td>
<td>172.1</td>
<td>270.8</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>735</td>
<td>41.2</td>
<td>123.9</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>712</td>
<td>40.5</td>
<td>184.7</td>
<td>139</td>
</tr>
<tr>
<td>Prismatic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blade</td>
<td>716</td>
<td>11.6</td>
<td>135.7</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>735</td>
<td>15.8</td>
<td>125.2</td>
<td>168</td>
</tr>
<tr>
<td></td>
<td>740</td>
<td>192.3</td>
<td>279.9</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>718</td>
<td>78.2</td>
<td>183.2</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td>734</td>
<td>78.8</td>
<td>143.1</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>747</td>
<td>72.1</td>
<td>145.8</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>695</td>
<td>76.7</td>
<td>167.3</td>
<td>132</td>
</tr>
<tr>
<td>Punch</td>
<td>726</td>
<td>85.35</td>
<td>54.13</td>
<td>156</td>
</tr>
<tr>
<td>Blade</td>
<td>778</td>
<td>63.35</td>
<td>82.84</td>
<td>214</td>
</tr>
<tr>
<td></td>
<td>748</td>
<td>101.62</td>
<td>51.54</td>
<td>175</td>
</tr>
<tr>
<td>Pressure</td>
<td>753</td>
<td>48.62</td>
<td>85.74</td>
<td>211</td>
</tr>
<tr>
<td>Blade</td>
<td>755</td>
<td>74.38</td>
<td>85.33</td>
<td>226</td>
</tr>
<tr>
<td></td>
<td>712</td>
<td>30.10</td>
<td>49.94</td>
<td>189</td>
</tr>
</tbody>
</table>
Throughout the experiments, reduction continued until the cores became exhausted and no more blanks could be removed, whether due to small core size, high platform angles, accumulated step or hinge terminations, or a combination of these factors. Although Eren et al. (2008) measured only formal blades and discoidal flakes while disregarding the products of core reshaping, here we define blanks as any removed flake larger than 2 cm. We adopted this size threshold, as flakes larger than 2 cm can easily be manipulated in the hand for tool use (Tomka 2001). Additionally, we sought to avoid complications arising from assuming knowledge about past knappers’ intentionality, particularly surrounding which removals they desired over which removals they considered waste. There is no a priori reason that prehistoric knappers would not have used the flakes produced from core reshaping. We are therefore measuring the maximum potential efficiency of each reduction sequence. This arbitrary threshold also allows for greater reproducibility of results, compared with methods that rely on subjective decisions regarding what constitutes a blank.

5.2.2 Reduction Sequences

In order to maintain experimental control, both the expert and intermediate knappers adhered to strict reduction sequences, reconstructed from archaeological, ethnographic and experimental sources. This section outlines the archaeological correlates of each of the eight reduction technologies and highlights the methodological aspects key to successfully accomplishing these technologies.

Originating in the Oldowan, but with perhaps even older roots at Lomekwi 3 in Kenya (Harmand et al. 2015), bipolar knapping is one of the oldest stone tool technologies and is executed by positioning a core on an anvil and striking the exposed platform until a flake is detached. Bipolar knapping proceeded in this relatively expedient fashion by exposing and striking new platforms until the cores were exhausted, following archaeological and ethnographic examples (de la Torre 2004; Diez-Martin et al. 2010; Jeske 1992; Martinez et al. 2010; Masao 1982; Semaw 2000; Shott 1989; Zaidner 2013).

Multiplatform reduction was conducted in this experiment via expedient and opportunistic selection of suitable platforms involving no constraints on the direction from which a flake can be removed. This sequence was reconstructed from a range of archaeological correlates (Brumm et al. 2010; Moore and Brumm 2007; Piperno et al. 2009; Robbins et al. 2000). With its origins in the Oldowan, the primary aim of multiplatform knapping is the production of as many large and usable flakes as possible, while not creating too high or too low edge angles that would inhibit further reduction.
Discoidal knapping involved the formation of a core with a bi-conical morphology, created via bifacial and radial flake removals. In order to maximise the use-life of the discoidal cores, both knappers intended each flake to both maintain this specific morphology as well as expose new suitable platforms. To maximise the utility and applicability of these results, the discoidal reduction sequences were modelled on well described reduction sequences (Boëda 1993; Bourguignon 1997; Delagnes and Meignen 2006; Eren et al. 2008), as well as archaeological examples from a range of regions and time periods (de la Torre 2004; de la Torre et al. 2003; Jaubert 1993; Jaubert and Farizy 1995; Locht and Swinnen 1994; Pasty 2000; Peresani 1998; Piperno et al. 2009).

With its roots in the Acheulean, bifacial knapping is an enduring and widespread technological innovation. However, in the interests of maintaining similar original nodule size and allowing the knapper to exploit the core until near exhaustion, as was the case with all other technologies, more recent and more heavily reduced bifacial technology is examined in this study. Reduction proceeded following archaeological examples of small foliate bifaces from the African Middle Stone Age (Archer et al. 2015; Shea 2008; Soriano et al. 2015; Villa et al. 2009) and the European Middle Palaeolithic (Rots 2009; Škrdla et al. 2014; Vaquero et al. 2001; Villa and Lenoir 2006; Zilhão 2009). Thin and invasive flakes were removed from both faces of the core, maintaining a sharp plane of intersection between the equivalent hemispheres.

Recurrent Levallois knapping, ubiquitous in the Middle Stone Age or Middle Palaeolithic, was conducted via establishing with radial flaking two asymmetrical hemispheres, one relatively flat upper hemisphere and one more protruding lower hemisphere. Meanwhile, the final platform was carefully faceted on the lower hemisphere. Following known reduction sequences (Boëda 1995; Bourguignon 1997; Brantingham and Kuhn 2001; Chazan 1997; Delagnes and Meignen 2006; Pelegrin 2009; Schlanger 1996; Van Peer 1992), and archaeological examples (Brantingham et al. 2000; Derevianko and Petrin 1995; Kuchikura and Watanabe 1973; Shipton et al. 2013a; Tryon 2006; Tryon et al. 2005; Usik and Demidenko 1993; Van Peer 1998; White and Ashton 2003), convexities were rigorously maintained on the upper surface in order to control the morphology of the recurrent Levallois flakes. These convexities were steepened or flattened with short dihedral flakes or invasive flakes respectively, with the intention of allowing the applied force to the faceted platform to remove a large portion of the upper surface without overshooting the core. This process of establishing two hemispheres and a faceted platform was repeated until no more recurrent Levallois flakes could be removed.
Prismatic blade core production in this experiment involved establishing a strong and flat, or slightly concave, platform from which to remove as many long and thin blades as possible. Following several archaeological examples (Delagnes 2000; Delagnes and Meignen 2006; Fisher 2006; Nishiaki 1989; Shimelmitz et al. 2011; Soriano et al. 2007; Wilkins 2012), blades were removed by striking the platform above a long and strong ridge on the core surface. Each successive blade removal created two new ridges at the intersection of flake scars, from which subsequent blades could be removed. Owing to the desire for long and thin flakes in blade reduction, overhang removal and abrasion is a particularly important aspect of this type of core reduction and was frequently conducted by the knappers. While this experiment involves unidirectional prismatic blade core knapping, bidirectional removals were at times used to maintain the core surface morphology or correct and straighten any haphazard ridges.

A variation of blade technology that occurred in the Upper Palaeolithic and onwards is the punch blade technique, in which one end of an intermediary tool, or ‘punch’, is placed on the core’s platform while the other end is struck by the percussor. This form of indirect percussion allows the knapper to situate the punch very close to the platform edge immediately above a ridge, thereby ensuring the precise placement of each blow. Reduction proceeded in this experiment using an antler punch and following experimental and archaeological examples from Mesoamerica, Europe and the Near East (Bordes and Crabtree 1969; Clark 2012; Combier 1967; Crabtree 1968; Golani 2013; Leduc 2012; Sheets 1977; Sørensen 2012; Tixier 1972).

Another blade technology of the Upper Palaeolithic and onwards is pressure blade manufacture, involving applying pressure from an indentor rather than using direct or indirect percussion. Like punch blade technology, the indentor can be very accurately placed, allowing greater control of blade production. The pressure blade component of this experiment involved a chest crutch and was conducted following extensive experimental and archaeological correlates (Desrosiers 2012; Flenniken and Hirth 2003; Pelegrin 2003; Sheets and Muto 1972; Sørensen 2012; Titmus and Clark 2003).

5.2.3 Flake Measurements
Due to the varied morphology of flakes, calliper measurements of cutting edge can be highly inaccurate. Therefore, cutting edge length was determined by measuring the outline of digital photographs of flakes placed ventral side down on a flat surface. Following the methodology of Eren et al. (2008), each blank larger than 2cm was photographed alongside a scale-bar using a digital camera. These images were imported to Adobe Photoshop CC and scaled to actual blank
size, then reduced to a polygon in Adobe Illustrator CC (Figure 5.1). This software was used to automatically trace the polygon’s perimeter and calculate the edge length in millimetres. Platforms and broken or dull edges were excluded from the perimeter measurement as they do not serve as a suitable cutting edge.

![Figure 5.1](image_url)

**Figure 5.1.** Demonstration of the method used to measure cutting edge length, showing a photograph of an original blank (a), and two stages in the process of reducing the photograph to a measurable polygon (b and c). Note the platform is excluded in the polygon measurement so as to measure possible cutting edge only.

Finally, in order to explore the possible reasons for any variation in the efficiency of the eight reduction sequences under examination, all complete and formal blanks were collected from each of the expert’s reduction sequences. These flakes were then weighed and measured using digital scales, callipers and a goniometer. The mass, dimensions (including length, mean width, mean thickness, platform width and bulb thickness), exterior platform angle (EPA), platform type, termination type, initiation type and platform preparation type were recorded for each flake. Flake thickness was assessed by averaging five thickness measurements taken at regular intervals on the flake. Bulb thickness was measured by subtracting the thickness of the flake at the apex of the bulb of percussion by the thickness of the flake immediately below the bulb of percussion, while accounting for any amorphous dorsal morphology. Finally, flake width was calculated by averaging the proximal, medial and dorsal width measurements. These measurements were taken to allow an exploration of the effects of flake size and shape on the cutting edge efficiency.
5.3 Core Reduction Efficiency
Throughout the 44 reduction sequences, a total of 30.40kg of flint was knapped, producing 5930 blanks with a cumulative cutting edge length of 613.53m. Table 5.2 shows the total values for each reduction strategy, summarising mass, count and cutting edge results, with a Kruskal-Wallis test for equal medians exploring the variability among the different reduction sequences.

Table 5.2. Mean mass, counts and cutting edge values for each reduction strategy. Kruskal-Wallis tests were conducted for each variable based on the values of the three or seven repetitions of each reduction method. Variables containing significant differences among the eight different technologies at the $\alpha = 0.05$ level are represented in bold. *The bipolar values were not included in the first five statistical comparisons, as significantly smaller cores were used owing to the typically small size of bipolar cores.

<table>
<thead>
<tr>
<th>Reduction Method</th>
<th>Mean initial nodule mass (g)*</th>
<th>Mean number of blanks*</th>
<th>Mean mass of all blanks (g)*</th>
<th>Mean mass of waste (g)*</th>
<th>Mean cutting edge (mm)*</th>
<th>Mean cutting edge per gram of core (mm/g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bipolar (N=3)</td>
<td>135.14</td>
<td>25.00</td>
<td>101.11</td>
<td>34.02</td>
<td>1404.33</td>
<td>10.36</td>
</tr>
<tr>
<td>Multiplatform (N=7)</td>
<td>723.86</td>
<td>103.43</td>
<td>565.28</td>
<td>158.60</td>
<td>12167.06</td>
<td>16.90</td>
</tr>
<tr>
<td>Discoidal (N=7)</td>
<td>734.86</td>
<td>106.14</td>
<td>545.17</td>
<td>189.69</td>
<td>12995.91</td>
<td>17.69</td>
</tr>
<tr>
<td>Biface (N=7)</td>
<td>734.29</td>
<td>191.14</td>
<td>576.45</td>
<td>158.05</td>
<td>14244.75</td>
<td>19.39</td>
</tr>
<tr>
<td>Levallois (N=7)</td>
<td>726.43</td>
<td>134.00</td>
<td>498.09</td>
<td>228.34</td>
<td>15857.12</td>
<td>21.83</td>
</tr>
<tr>
<td>Prismatic Blade (N=7)</td>
<td>726.43</td>
<td>134.43</td>
<td>482.76</td>
<td>243.67</td>
<td>16303.79</td>
<td>22.46</td>
</tr>
<tr>
<td>Punch Blade (N=3)</td>
<td>750.67</td>
<td>181.67</td>
<td>604.38</td>
<td>146.28</td>
<td>17236.65</td>
<td>22.95</td>
</tr>
<tr>
<td>Pressure Blade (N=3)</td>
<td>740.00</td>
<td>208.67</td>
<td>615.24</td>
<td>124.70</td>
<td>18875.14</td>
<td>25.91</td>
</tr>
</tbody>
</table>

While significant differences occur among the blank counts ($H = 28.69$, d.f. = 6, 34, $p < 0.001$), Mann-Whitney tests with Bonferroni corrections (counteracting the increased risk of a type-I error during multiple comparisons) reveal that, excluding the bipolar reductions, the only significant difference is that the biface reduction sequences produced significantly more blanks on average than the multiplatform and discoidal repetitions ($U = 24.5$, $p = 0.045$ for both). This discrepancy is likely explained by the typically higher fragmentation rates that accompanies biface reduction.

Bifacial knapping involves the concerted production of very thin, expanding flakes called thinning flakes, which increases the likelihood of breakage. This means however, that the number of flakes produced is unlikely to be an adequate representation of raw material efficiency. Instead we turn to the length of cutting edge produced per gram of core to assess raw material efficiency among the sample of eight knapping technologies. Table 5.2 shows the total cutting edge per gram for each reduction strategy, with an ascending trend through this order of reduction sequences. Figure 5.2 explores this pattern further, by plotting the total cutting edge per gram for each repetition.
A Kruskal-Wallis test reveals that significant differences occur among the different reduction strategies (\(H = 22.92, \text{ d.f.} = 7, 36, p = 0.0018\)), however the subsequent Mann-Whitney pairwise post-hoc analysis with Bonferroni corrected p-values returned no significant results. This means that no individual knapping strategy was significantly more efficient than another. With the original Kruskal-Wallis test suggesting that significant differences do occur among the samples, the variability among the eight different technologies was examined further by combining each technology into broad time periods. These were the Oldowan (bipolar, multiplatform and discoidal), Middle Palaeolithic (biface, Levallois and prismatic blade) and Upper Palaeolithic and onwards (punch blade and pressure blade) (Figure 5.3). Lithic technologies are not produced in isolation of course. Prehistoric toolkits would have consisted of varying proportions of the available technologies at the time, depending on raw material availability and prospective function. Therefore, combining these eight technologies into broad time periods will allow for more meaningful comparisons of broad-scale temporal trends in tool-kit efficiency. Ascribing these
technologies to different time periods is done cautiously however, as our knowledge of evolution in lithic technology is being constantly revised. If certain technologies receive older or younger dates, then the following analysis could easily be updated to reflect any changes.

**Figure 5.3.** Bar chart, with one standard error bars and each data point superimposed, of the eight technologies grouped into their corresponding time periods, showing the Oldowan, consisting of bipolar, multiplatform and discoidal technologies (N = 17, μ = 16.07), the Middle Palaeolithic, consisting of biface, Levallois and prismatic blade technologies (N = 21, μ = 21.23), and the Upper Palaeolithic and onwards, consisting of punch blade and pressure blade (N = 6, μ = 24.22).

Significant differences occur among these three grouped samples (H = 17.13, d.f. = 2, 41, p < 0.001), with a Mann-Whitney pairwise test with Bonferroni corrected p-values revealing that the Middle Palaeolithic and Upper Palaeolithic reduction sequences produced a significantly greater length of cutting edge per gram of original core compared with the Oldowan technologies (U = 64, p = 0.0025 and U = 6, p = 0.0055 respectively). Despite a difference of more than 3mm/g of cutting edge length per gram of core between the Middle and Upper Palaeolithic and onwards technologies, this difference is not significant (U = 30, p = 0.17). These results reveal that the transition from Lower to Middle Palaeolithic toolkits was accompanied by an increase in the efficiency of cutting edge production per mass of core. On the other hand there appears to be no inherent increase in the core efficiency at the transition from the Middle to Upper Palaeolithic.
Interestingly, pressure blades outperformed all other core technologies tested in this experiment. While the technologies examined here that formed a component of the Upper Palaeolithic and onwards are by no means significantly more efficient than the preceding period, it would appear that the evolution of cutting edge efficiency that is evident by the ascending trend in Figure 5.2 continued during the Upper Palaeolithic, Epipalaeolithic, Mesolithic and Neolithic.

### 5.4 Causes of Variability in Efficiency

Having examined the broad temporal trend in cutting edge efficiency, we now turn to the individual flake attributes which contribute to this variability. Based on measurements from the sample (N = 488) of complete and formal flakes produced by the 19 experimental reduction sequences conducted by the expert knapper, we can identify features of flakes which maximise flake economy. Figure 5.4 plots cutting edge length per gram against nine flake attributes, most of which reveal power relationships between the axes. To present these trends more clearly, both axes for all nine charts were transformed to linear relationships using the natural log (ln). From these charts, it is clear that minimising flake mass (R² = 0.898), flake thickness (R² = 0.935), bulb thickness (R² = 0.462), flake width (R² = 0.727), platform depth (R² = 0.611) and platform width (R² = 0.557) all strongly contribute to maximising the cutting edge length per gram of individual flakes. These results partly confirm the findings of Lin et al. (2013), who found reducing flake thickness, bulb thickness and platform size had a positive effect on flake economy.
Figure 5.4. Scatter plots with both axes transformed using the natural log (ln) examining the influence of mass, thickness, bulb thickness, length, width, elongation, platform depth, platform width and exterior platform angle (EPA) on the cutting edge length per gram of core for individual flakes. The sample size of each scatter plot is 488, except for the platform depth, platform width and EPA scatter plots, which had sample sizes of 460 owing to the presence of some crushed platforms.

Where our findings diverge is in the role of flake size, elongation (length divided by thickness), and EPA. Lin et al. (2013) used geometric models and flake measurements to hypothesise that increasing flake size, the ratio of length to width, and EPA should maximise the economy of flakes.
Our findings suggest, however, that minimising flake size (mass, width and thickness) is key to maximising cutting edge length per gram. Moreover, flake length ($R^2 = 0.168$) or elongation ($R^2 = 0.136$) had very little influence on the raw material efficiency of the flakes. Increasing length relative to width had only a very weak impact on cutting edge length per gram, which was far superseded by other size attributes like minimising thickness and width. It is interesting therefore that the three blade technologies, all of which maximise elongation, were the most efficient at cutting edge production. This could largely be credited to the production of narrow and thin blades within these knapping schemas, rather than the elongate nature of blades. The weakly positive relationship between elongation and cutting edge per gram is likely explained by what Lin et al. (2013) identify as the ‘square cube principle of proportional solids’, whereby increases in the surface area of an elongate flake results in a lesser increase in volume compared with a more circular flake. Lastly, EPA, which was identified as a fundamental component of flake economy by Lin et al. (2013), appears in this present study to have a negligible ($R^2 = 0.0096$) impact on efficiency in terms of cutting edge per gram of original core.

In terms of the qualitative features of flakes, Figure 5.5 shows boxplots of the cutting edge length per gram of each flake according to different platform, termination and platform preparation types. A Kruskal-Wallis test for equal medians reveals that platform type has a significant impact on the production of cutting edge length per gram of core ($H = 125.5$, $p < 0.001$). Mann-Whitney tests with Bonferroni corrections reveal that flakes with focalised platforms have significantly greater cutting edge length per gram than those with dihedral ($U = 2080$, $p < 0.001$) or plain ($U = 14820$, $p < 0.001$) platforms. Termination type also significantly influences cutting edge efficiency ($H = 22.9$, $p < 0.001$), with feather terminations facilitating higher cutting edge per gram of core than plunging ($U = 2517$, $p = 0.002$), or step or hinge ($U = 8426$, $p = 0.001$) terminations. Finally, platform preparation strategies are similarly effective at increasing cutting edge efficiency ($H = 74.22$, $p < 0.001$), with the use of either overhang removal or faceting resulting in significantly higher cutting edge per gram than flakes without platform preparation ($U = 8803$, $p = 0.044$). Additionally, flakes exhibiting both overhang removal and faceting performed significantly better than those without preparation ($U = 4627$, $p < 0.001$), as well as those with only one type of preparation ($U = 10550$, $p < 0.001$). Attributes like platform, termination and platform preparation type are all associated with the quantitative associations above. For example, focalised platforms, feather terminations and extensively prepared platforms all contribute to producing flakes with low thickness, bulb thickness and flake width values.
These analyses offer a holistic identification of the features that make a flake efficient in terms of the cutting edge length produced per gram of original core. In summary, it appears that the efficiency of flakes are negatively impacted by areas of mass on a flake that do not contribute to the cutting edge, such as a bulb or amorphous dorsal surface, as well as portions of the flake perimeter that do not contribute to the cutting edge, such as platforms or steep broken edges. The most efficient flakes, therefore, are those that are thin and narrow, with diffuse bulbs, small platforms, feather terminations and extensive platform preparation.

### 5.5 Skill

Finally, we seek to examine the role of knapping skill on the cutting edge length per gram efficiency of each reduction sequence. As mentioned in the methods section, bipolar reduction was conducted by the expert knapper only due to the extremely low skill required, and punch and pressure blade reduction was conducted by the expert knapper only owing to the high level of skill required. In this section therefore, we examine the influence of skill level on the cutting edge efficiency of multiplatform, discoidal, biface, Levallois and prismatic blade core reduction (Table 5.3).
<table>
<thead>
<tr>
<th>Reduction Method</th>
<th>Skill Level</th>
<th>Mean number of blanks</th>
<th>Mean mass of waste (g)</th>
<th>Mean cutting edge per gram (mm/g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiplatform</td>
<td>Expert</td>
<td>67</td>
<td>62.5</td>
<td>12.43</td>
</tr>
<tr>
<td></td>
<td>Intermediate</td>
<td>118</td>
<td>197.04</td>
<td>18.69</td>
</tr>
<tr>
<td>Discoidal</td>
<td>Expert</td>
<td>110</td>
<td>95.6</td>
<td>18.23</td>
</tr>
<tr>
<td></td>
<td>Intermediate</td>
<td>104.6</td>
<td>227.32</td>
<td>17.48</td>
</tr>
<tr>
<td>Biface</td>
<td>Expert</td>
<td>223</td>
<td>202.71</td>
<td>21.06</td>
</tr>
<tr>
<td></td>
<td>Intermediate</td>
<td>178.4</td>
<td>140.19</td>
<td>18.73</td>
</tr>
<tr>
<td>Levallois</td>
<td>Expert</td>
<td>140.5</td>
<td>145.55</td>
<td>22.68</td>
</tr>
<tr>
<td></td>
<td>Intermediate</td>
<td>131.4</td>
<td>261.46</td>
<td>21.50</td>
</tr>
<tr>
<td>Prismatic Blade</td>
<td>Expert</td>
<td>156</td>
<td>144.15</td>
<td>26.09</td>
</tr>
<tr>
<td></td>
<td>Intermediate</td>
<td>125.8</td>
<td>283.48</td>
<td>21.01</td>
</tr>
</tbody>
</table>

In all bar one reduction sequence (biface), the expert knapper produced much less waste debitage compared with the intermediate knapper. However, this represents only a relative assessment of raw material efficiency. Therefore, two-sample t-tests were used to explore any significant variation between the cutting edge per gram output of the expert and intermediate knappers for each of the four reduction strategies. There was no significant difference between the cutting edge per gram efficiency of multiplatform (t = -1.86, d.f. = 1, 5, p = 0.12), discoidal (t = 0.25, d.f. = 1, 5, p = 0.82), biface (t = 1.12, d.f. = 1, 5, p = 0.31) or Levallois (t = 0.43, d.f. = 1, 5, p = 0.69) knapping. Comparatively, the expert knapper produced significantly more cutting edge per gram for the blade core iterations compared with the intermediate knapper (t = 4.76, d.f. = 1, 5, p = 0.005).

This does not mean that the intermediate knapper necessarily executed the multiplatform, discoidal, biface and Levallois reduction strategies as effectively as the expert knapper however. For example, general observations of Levallois flake size and shape as well as the number of recurrent Levallois flakes successfully removed suggest that the expert knapper more effectively performed the Levallois experiments. What can be concluded is that the raw material efficiency of multiplatform, discoidal, biface and Levallois technology is less sensitive to reductions in knapper skill compared to prismatic blade technology. In other words, equivalent cutting edge is produced from these knapping strategies regardless of whether a less skilful knapper imperfectly executes the reduction sequence and produces less technologically typical flakes, such as broad dihedral flakes for discoidal knapping, or thin and large Levallois flakes. It is possible that prismatic blade knapping was more efficiently performed by the expert knapper because the desire for thin and long flakes in this technology increases the likelihood of snaps and hinge or step terminations, the correction of which can waste valuable raw material. The elongate core face typical of prismatic blade reduction...
also makes remedying such mistakes more difficult and costly in terms of raw material usage, as these mistakes tend to be further from the platform.

5.6 Discussion and Conclusions
This study investigated the raw material efficiency of eight different lithic core technologies by measuring the ratio of cutting edge length to original core mass. The results garnered from the 5930 blanks produced in the experiments revealed a gradual upward trend in raw material efficiency through the sequence of bipolar, multiplatform, discoidal, biface, Levallois, prismatic blade, punch blade and pressure blade technologies (Figure 5.2). Interestingly, no statistically significant differences occurred among the individual reduction strategies. Any changes in raw material efficiency occurring throughout the evolution of stone tool technology therefore appear to be gradual. These changes were only perceptible when viewing prehistoric tool-kits on a broader-scale, by grouping each technology into their broad time periods. This revealed a significant difference between the raw material efficiency of the technologies typically made in the Lower Palaeolithic and those typically made in the Middle Palaeolithic. In contrast, no significant difference occurred among the Middle Palaeolithic technologies and those in this sample that were made in the Upper Palaeolithic, Epipalaeolithic, Mesolithic and Neolithic.

The fact that the cutting edge lengths per gram of all eight technologies were statistically indistinguishable highlights the shortfalls of comparing only two lithic technologies at a time as was done in all previous comparisons. An experimental comparison of two technologies is likely to confirm the null hypothesis that no significant difference in raw material efficiency exists. By examining eight technologies which broadly span the evolution of lithic technology from the Oldowan to the Neolithic, we identified statistically significant trends in raw material efficiency over time. The null hypothesis, that no significant differences in cutting edge per gram of core occur among the eight examined technologies, can therefore be rejected as the technologies ascribed to the Middle Palaeolithic were more efficient than those ascribed to the Lower Palaeolithic.

While Eren et al. (2008) sought to examine the Middle to Upper Palaeolithic transition using discoidal and prismatic blade core technology, more recent dates of prismatic blades situate their emergence long before the Upper Palaeolithic (Johnson and McBrearty 2010; Shimelmitz et al. 2011; Soriano et al. 2007; Wilkins 2012). Meanwhile, discoidal technology is better situated in the Lower Palaeolithic (de la Torre 2004; de la Torre et al. 2003; Piperno et al. 2009). Therefore, while they found prismatic blades to be no more efficient than discoidal flakes, what was really being compared was the Lower and Middle Palaeolithic. We can therefore, for the first time, conclude that
it is unlikely that the Middle to Upper Palaeolithic transition was accompanied by an increase in the raw material efficiency of the available toolkits.

It should of course be noted that there are far more than eight lithic technologies, but with all other comparisons of core efficiency comparing no more than two technologies, we offer a step in the right direction. By selecting representative technologies from different periods, we aimed to capture much of the variation occurring over the sweep of human evolution. Ascribing certain lithic technologies to certain time periods, as we have attempted here, is a difficult task as the picture of evolution in lithic technology becomes increasingly branching and multidirectional. This was done in order to provide a broad-scale picture of changes in efficiency, but should be considered with caution as new sites and dates arise. We hope this present study offers a broad and exploratory assessment that could be used as a platform for more focussed and site-specific comparisons of technological efficiency.

We additionally sought to identify attributes of individual flakes which maximise their ratio of cutting edge to flake mass. The measurements and qualitative attributes of 488 complete flakes revealed that the most efficient flakes are those that are small, thin and narrow, with diffuse bulbs, small platforms, feather terminations and extensive platform preparation. Interestingly, elongation and exterior platform angle had minimal to negligible effects on raw material efficiency. It is therefore no surprise that the pressure blade cores performed the most efficiently of all eight technologies under investigation, as pressure blade manufacture involves taking the notion of platform preparation and isolation, key factors in minimising flake thickness, width and bulb thickness, to the extreme. These findings also have significant implications for assemblages with flakes possessing these optimal attributes. Microblade and microlithic technologies, sometimes made via the pressure technique, typically possess these traits and may therefore represent an optimisation of lithic technology geared towards maximising efficiency, whether a conscious attempt or a persistent behavioural adaptation. For example, microliths have been linked to periods of environmental, demographic or social stress, making such technologies likely strategies for offsetting risk in scenarios of raw material scarcity or environmental stress (Hiscock et al. 2011). Further research is required to investigate this possible association between lithic raw material scarcity and strategies which optimise the cutting edge efficiency of flakes.

These results suggest that throughout our biological and cognitive evolution, the major evolution in cutting edge efficiency likely occurred around the transition from the Lower Palaeolithic to the Middle Palaeolithic. The transition from the Middle to Upper Palaeolithic on the other hand, does
not appear to be accompanied by a toolkit-wide increase in raw material efficiency. This means that the toolkits of the Neanderthals and their contemporaneous Homo sapiens exhibited comparable degrees of raw material efficiency. However, we demonstrated that pressure blade technology involved the highest raw material efficiency of the eight technologies investigated. Therefore after this transition, during parts of the Upper Palaeolithic, Epipalaeolithic, Mesolithic and Neolithic, Homo sapiens continued developing their blade core technology to produce more efficient blank technologies. Minimising flake thickness, bulb thickness and flake width was achieved via specialised blade knapping techniques like pressure knapping, rather than direct percussion. While this technique requires greater investments in preparatory time, through pressure indentor manufacture as well as more intensive platform preparation, it allows for heightened raw material efficiency. Future research is needed to investigate the relationship between heightened investment and raw material efficiency, and model whether these strategies represent an optimisation of the knapping process.
CHAPTER 6:
Measuring behavioural and cognitive complexity in lithic technology throughout human evolution

Muller, A., C. Clarkson and C. Shipton (submitted) Measuring behavioural and cognitive complexity in lithic technology throughout human evolution.

<table>
<thead>
<tr>
<th>Contributor</th>
<th>Statement of contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antoine Muller (Candidate)</td>
<td>Experimental design (50%)</td>
</tr>
<tr>
<td></td>
<td>Analysis (80%)</td>
</tr>
<tr>
<td></td>
<td>Writing (80%)</td>
</tr>
<tr>
<td></td>
<td>Editing (80%)</td>
</tr>
<tr>
<td>Chris Clarkson</td>
<td>Experimental design (30%)</td>
</tr>
<tr>
<td></td>
<td>Experimental sample (100%)</td>
</tr>
<tr>
<td></td>
<td>Analysis (10%)</td>
</tr>
<tr>
<td></td>
<td>Writing (10%)</td>
</tr>
<tr>
<td></td>
<td>Editing (10%)</td>
</tr>
<tr>
<td>Ceri Shipton</td>
<td>Experimental design (20%)</td>
</tr>
<tr>
<td></td>
<td>Analysis (10%)</td>
</tr>
<tr>
<td></td>
<td>Writing (10%)</td>
</tr>
<tr>
<td></td>
<td>Editing (10%)</td>
</tr>
</tbody>
</table>

Abstract
Stone tool making, observed archaeologically from 3.3mya, involves complex problem solving and forethought, but the relative complexity of different Palaeolithic technologies remains unknown. Decision making in replicative knapping is here used to explore the degree of behavioural and cognitive complexity involved in five different types of stone tool manufacture (bipolar, discoidal, biface, Levallois and prismatic blade) that represent the evolution of core reduction strategies from the Oldowan through to the Upper Palaeolithic. While some hypothesise that each key transition was marked by an increase in cognitive complexity, such hypotheses remain untested assumptions. Determining the level of behavioural complexity involved in each of these core reduction strategies using problem-solution distance modelling offers a means of detecting significant increases in the level of human cognitive complexity displayed over time. To directly test for differences in
complexity among knapping schema, replication experiments were conducted by two highly skilled knappers. Experiments were filmed and the duration of different stages in the sequence was annotated. Hierarchical diagrams were produced showing the organisation of the different actions involved in stone tool knapping. The results show a pattern of increasingly complex behaviour through the sequence of bipolar, discoidal, biface, prismatic blade and Levallois knapping. Furthermore, Neanderthals and their contemporaries, *Homo sapiens*, both employed knapping technologies exhibiting comparably high levels of complexity.

6.1 Introduction

Recognising cognitive and behavioural complexity in the archaeological record is a key concern in palaeoanthropological research, typically with a focus on such traits as personal adornment, art, pigments, complex burials, carved bone, long distance exchange, use of adhesives, and complex projectiles (Ambrose 2001; 2010; Cain 2006; Chase and Dibble 1987; Clark 1989; Deacon 1989; Foley and Lahr 1997; Henshilwood et al. 2009; Henshilwood and Marean 2003; Klein 1995; Langley et al. 2008; Marean et al. 2007; McBrearty and Brooks 2000; Mellars 1989a; 1989b; 1991; Milo 1998; Thackeray 1992; Wadley 2010). Such items are rare in the archaeological record however, hampering a geographically comprehensive and temporally deep perspective on cognition. The ubiquity of stone tools in the Palaeolithic record offers an alternative means of assessing cognitive complexity, but quantifying and comparing lithic behavioural complexity remains an elusive goal. This study investigates behaviour and cognition by examining the decision making involved in different types of stone tool manufacture, including bipolar, discoidal, biface, Levallois and prismatic blade core knapping (Figure 6.1). These technologies broadly represent the evolution of lithic core technology from the Oldowan (c. 2.5mya) to the Upper Palaeolithic (c. 50kya).
Behavioural complexity is modelled in this paper via observations and analysis of footage of replicative knapping experiments conducted by two expert knappers familiar with each reduction strategy. By identifying the minutia of stages involved in these knapping sequences, and analysing these stages through the lens of concepts borrowed from neuroscience and psychology (discussed further below), the complexity of the decision making processes involved in the manufacture of different stone tools can be reconstructed. Behaviour is defined for the purposes of this paper as the actions undertaken by hominins, while cognition refers to the underlying brain functions that facilitate those actions.
To assess the behavioural complexity of different knapping tasks we use problem-solution distance modelling. Problem-solution distance modelling is based on the idea that ‘indirect thinking’, developed by Köhler (1925), plays a crucial role in complex tasks such as tool making. Indirect thinking involves setting aside an immediate desire or problem in favour of addressing a series of intermediate goals, the completion of which culminate in the satisfaction of the original desire (Köhler 192511). This approach, pioneered in archaeology by Haidle and Lombard (Haidle 2009; 2010; 2012; Lombard 2012; Lombard and Haidle 2012) is a means of reconstructing the number and nature of steps involved in complex technological tasks. The problem-solution distance is a measure of the conceptual distance between the perception of the original problem or need, and the final satisfaction of that need (Haidle 2010). This method requires identifying and sequencing each step involved in overcoming a problem, meaning that the greater the number of steps, the greater the problem-solution distance. Simple knapping techniques with short operational sequences would therefore have short problem-solution distances and require less complex behaviour and cognition compared with more complex lithic technologies.

A key factor which heavily influences the magnitude of the problem-solution distance is the degree of hierarchical organisation involved in a process. Hierarchical organisation is the extent to which different component parts must take place in a particular order to achieve an overarching goal (Greenfield 1991; Holloway 1969; Stout 2011). In hierarchical sequences, broad higher-order actions are divided into subordinate actions. The fulfilment of a set of subordinate tasks allows each consecutive superordinate task to be achieved until the overarching goal is accomplished. In Levallois technology for example, the faceting of a striking platform is a subordinate task undertaken in order to achieve the overarching goal of creating a strong, high angled platform from which to strike a large flake from a carefully shaped core face (Boëda 1995; Shipton et al. 2013b).

6.1.1 Hypothesis Testing

Based on existing archaeological and experimental studies, it is possible to hypothesise the relative problem-solution distances of the different knapping strategies under investigation in the present study. For example, discoidal knapping involves the simultaneously completed tasks of removing flakes and exposing new exploitable platforms from the negative scar left by the removed flake (de la Torre 2004). This leads to the hypothesis that discoidal knapping will require greater behavioural complexity than bipolar knapping, a technology that involves a more expedient system of flake removals. In terms of Acheulean biface technology, much literature has highlighted the symmetry and seeming imposition of form involved (Cole 2015; Goren-Inbar 2011; Gowlett 1988; 2011; Machin et al. 2007; Shipton et al. 2013b; Stout 2002; Stout et al. 2008; Wynn 2002), leading to the
hypothesis that Acheulean biface technology will be more behavioural complex than the preceding
technologies. Meanwhile, the hierarchical tasks involved in Levallois knapping (establishing the
upper hemisphere, lower hemisphere and final platform) and prismatic blade knapping (platform
and core face maintenance) result in the hypothesis that these technologies should exhibit greater
behavioural complexity than the other three under investigation (Ambrose 2001; Gowlett 1984;
Moore 2010; Shipton et al. 2013b; White et al. 2011; Wynn and Coolidge 2010a). The relative
behavioural complexity between Levallois and prismatic blade knapping has been the topic of much
debate, with several assumptions and hypotheses remaining thus far untested. For example, Bar-
Yosef and Kuhn (1999) hypothesise that the complexity and cognitive implications of blade
technology have been largely overstated. This present study aims to quantitatively compare the
behavioural complexity involved in these five technologies, thereby testing the assumptions and
hypotheses outlined above. Without quantitative testing, these pervasive assumptions are at risk of
becoming unfounded truisms.

6.2 Materials and Methods

6.2.1 Reduction Sequences

The experimental reduction sequences and their archaeological correlates for bipolar, discoidal,
biface, Levallois and blade manufacture followed here are described in a recent study (Muller and
Clarkson 2016a). Any additional instructions given to the expert knappers or deviations from this
method are described here.

The bifaces produced in the associated study (Muller and Clarkson 2016a) were modelled on
relatively recent, small foliate pieces. In this present study, the larger bifaces of the Acheulean
(especially handaxes) were the subject of investigation. In order to represent the vast temporal span
of handaxe manufacture while also capturing the upper extent of cognitive potential involved, three
of the handaxes were highly formalised via intensive shaping and thinning through raising the plane
of intersection and preparing platforms to strike more invasively (Late Acheulean). For the other
two handaxes the knapper did not use these techniques, resulting in pieces that were more roughly
finished with remnant cortex and sinusoidal edges (Early Acheulean). Otherwise, biface knapping
in this experiment involved forming a rounded base and a pointed or ovate tip with an axis of
symmetry from the tip to the centre of the base.

In order to capture the extent of variability involved in Levallois knapping, a variety of iterations
were conducted in the experiments, including preferential, recurrent preferential, and Levallois
point sequences (see Boëda 1995; Delagnes and Meignen 2006). Recurrent centripetal Levallois
flaking was not conducted in the experiment, but it is anticipated that this method would be somewhat less complex due to the more expedient selection of platforms from which to remove Levallois flakes. When necessary, a *chapeau de gendarme* striking platform, or slight protrusion at the centre of the platform, was formed by the knapper in order to guide the preferential strike to a precise point and remove the desired flake (see Pelegrin 2009).

Both unidirectional and bidirectional percussive blade technologies were employed in this study. For the unidirectional iterations the majority of removals were initiated from a primary platform, however a flexible process of bidirectional core maintenance was permitted. To account for some additional variability in blade knapping, some of the iterations involved the maintenance of parallel ridges via cresting.

These archaeological and experimental descriptions of the five technologies allowed the knapper to follow well established reduction sequences with enough fidelity to be under similar constraints as prehistoric knappers, while allowing a suitable level of freedom to react to the changing morphologies of the cores. Additionally, incorporating several variants of some of the technologies under investigation serves to test the reliability of this method, with more variation in hierarchical complexity expected between than within the five technologies under investigation.

Additionally, this research models the behavioural complexity required for a specific knapping technology, not the underlying cognition itself. As not all behaviours require an individual’s most sophisticated cognition, archaeologists can only ever model the ‘minimum necessary competence’ (Wynn 1985). In other words, our aim was simply to document the level of cognitive complexity needed to complete the task, not the maximum cognitive capabilities of the knapper’s themselves. In this way, the cognition needed for different stone tool technologies can be estimated, despite different technologies being made by different hominin species. In order for this minimum necessary competence to relate to cognitive capacities as closely as possible, the most complex core reduction strategies from each technological period were chosen to replicate, so that the observable behavioural complexity approaches the limits of hominin cognitive complexity.

### 6.2.2 Knapping Experiments

The filmed knapping sequences were performed by two expert knappers. The results extracted from the footage of the first expert knapper (CC) forms the mainstay of the results, while the footage of the second expert (Jacques Pelegrin) is primarily used to confirm that any patterns observed from the first knapper are not strongly influenced by an individual’s knapping technique. This second
knapper was not shown the footage of the previous experiments. To further limit the role of individual knapper variation, both knappers closely adhered to known knapping sequences and their archaeological correlates. Despite this experimental control, a degree of decision making is unavoidably left to the discretion of the experimental knappers. This raises the issue of modern knappers having a limited range of resources from which to learn, meaning that modern knappers may possess similar solutions to similar problems that may not reflect the solutions of past knappers. Many current knappers followed similar learning pathways, with most acquiring expertise from instructions from a limited range of experts and literature. Finding knappers from different learning pathways therefore becomes an important component of experiments concerning decision making. Importantly, both knappers gained their expertise independently. One was largely self-taught from a range of videos and literature, whereas the other was taught via instruction from other experts. Despite relying on diverse experimental knappers, the problem of differences between the decision making processes of past and present knappers is likely an inescapable one. However, this problem can be increasingly mitigated as more archaeological evidence of the decision making involved in reduction sequences is compiled.

For the first expert knapper, a total of 17 experiments were filmed; two for bipolar, two for discoidal, four for biface, five for Levallois and four for blade core knapping. A larger number of iterations were performed on the biface, Levallois and blade reduction strategies in order to better capture the varieties of these technologies. These experiments produced more than four hours of knapping footage (footage of examples of entire reduction sequences are available at https://www.youtube.com/user/clarchaeology). To explore the role, if any, of individual knapping variability, the second expert knapped four reduction sequences; one discoidal, one Late Acheulean biface, one recurrent preferential Levallois and one bidirectional blade core. Only the first expert knapper conducted the bipolar iterations, as this technology involves such little complexity that almost no variation between knappers was expected.

Due to personal preference one expert knapper employed copper headed billets, while the other expert used a combination of soft stone, antler and wood billets. Copper headed billets serve as suitable analogues to soft stone or antler in terms of mass and hardness (Clark 2012; Crabtree 1967; 1968; Sheets and Muto 1972), while also having the advantage of being symmetrical, standardised in shape and durable enough to last for the entire experiment. It was also recently demonstrated that the morphology of flakes derived from copper billets are statistically indistinguishable from common natural soft billets such as antler and soft stone (Muller and Clarkson 2016a Supporting Information). The potential for variability caused by the knappers using different billets was
deemed to be a lesser problem than instructing them to use billets with which they are less familiar. In order to avoid raw material variability, cobbles of the same high quality cryptocrystalline flint were used throughout the experiment.

Once the footage was compiled, analysis involved recording each individual action, such as core rotations, core examinations, overhang removal and knapping strikes, as well as the duration of each action, timed to the nearest tenth of a second using the video player timestamp. From consultation with the expert knappers, as well as known archaeological and experimental reduction sequences (described above), these actions were grouped into stages. The duration of each stage was calculated by combining the duration of all the actions within that stage. These stages (referred to in this experiment as ‘sub-foci’) were then organised into hierarchical diagrams, with each stage forming a discrete component of the hierarchical diagram.

Behaviour and cognition are enigmatic concepts, so providing means of testing the role of confounding variables is fundamental to this study. The major potential source of confounding influence is the concept of equifinality, the phenomenon where the same end goal can be achieved via multiple pathways. Different knappers may achieve the same lithic end-product via different means. For example, one knapper may prefer to establish the upper surface of a Levallois core prior to the lower surface, and another may tend to conduct this sequence in reverse. We argue, however, that there are bottlenecks in the hierarchical organisation of lithic technology that prevent substantial variation in structure. In other words, there are hierarchically ordered actions that must be completed for certain lithic technologies regardless of an individual’s approach to that technology. For instance, Levallois technology involves three hierarchically ordered tasks; establishing the lower surface, establishing the upper surface and preparing a suitable final platform. The qualitative and quantitative methods described below are designed to capture such hierarchical organisation. Regardless of who is knapping the Levallois core, these three goals must be achieved. The order of the completion of these tasks, the hammers used, and whether these tasks can be completed in tens of strikes or hundreds of strikes is irrelevant to the hierarchical complexity involved. The metrics for estimating hierarchical complexity outlined below are designed to be insensitive to these idiosyncrasies of individual knappers. Instead, they were designed to only capture the degree of hierarchical organisation. The second knapper was incorporated in this study in order to test whether these methods are adequately insensitive to individual knapper variation. If there is more variation between the expert knappers than among different lithic technologies, then this method can be deemed invalid. However, if little to no variation exists between the two
knappers for each technology, but variation exists between the five different technologies, then we can be confident that this method is measuring actual differences in hierarchical complexity.

6.2.3 Hierarchical Diagrams

The hierarchical diagrams were based on the work of Haidle and Lombard (Haidle 2009; 2010; 2012; Lombard 2012; Lombard and Haidle 2012) regarding non-hominin tool use. As of yet, these diagrams offer the only real replicable means of comprehensively visualising and quantifying complex cognitive tasks. To tailor this method to the investigation of knapping reduction sequences, these diagrams have been revised by incorporating aspects of other hierarchical cognitive models (for example Greenfield 1991; Greenfield and Schneider 1977; Moore 2003; 2007; 2010; 2011; 2013; Stout 2011; Stout et al. 2008).

The incorporation of hierarchical complexity separates this study from many of the previous attempts at modelling complexity in stone tool manufacture. For example, Perreault et al. (2013) made a significant advancement in our understanding of technological complexity by counting the number of ‘procedural units’ involved in knapping, a remodelling of Oswalt’s (1976) ‘techno-units’. However, Perreault et al. (2013) only tally technological complexity, whereas the present study ranks these steps hierarchically and examines when and how often steps are repeated.

A notable study involving a hierarchical reconstruction of the steps of lithic production was conducted by Mahaney (2014), who explored the common elements of hierarchical behaviour between handaxe manufacture and the evolution of syntactic language. This study however, explored only one type of lithic production and was conducted using linguistic models. The present study compares five types of lithic reduction and is grounded in stone tool analysis.

The result of consolidating these different models was a series of hierarchical diagrams, depicting a ‘primary focus’ that is divided into a series of branching and hierarchically structured stages or ‘sub-foci’ (adapted from Haidle 2009; 2010; 2012; Lombard 2012; Lombard and Haidle 2012). A primary focus can be defined as the original problem that requires attention. In each of the hierarchical diagrams in the results section, the primary focus is the creation of sharp edged tools. Meanwhile, sub-foci are simply secondary problems on which a knapper is concentrating. A group of sub-foci on the same hierarchical level are grouped into chains of sub-foci, or ‘phases’, connected by the arrows in the diagrams. By alternating their attention between these different phases, a knapper may achieve the primary focus.
Figure 6.2 defines each component of the hierarchical diagrams. These diagrams represent an aggregation of all these components, thereby showing the activities on which a knapper must focus. They outline the sub-goals of the knapping process that must each be completed in order to achieve the primary goal. The hierarchical modelling undertaken in this present study differs from preceding attempts at examining hierarchical complexity (for example Greenfield 1991; Greenfield and Schneider 1977; Moore 2003; 2007; 2010; 2011; 2013; Stout 2011; Stout et al. 2008), in that high-resolution data were collected for the duration of each specific sub-focus involved in the reduction sequences.

**Figure 6.2.** Definitions of each component of the hierarchical diagrams.

The hierarchical diagrams were constructed by identifying each individual sub-focus enacted throughout the sequence. For example, in biface knapping (Figure 6.5), the diagram was constructed by identifying the first sub-focus, ‘produce biface’, and considering where attention must be directed in order to complete this sub-focus. The sub-focus ‘establish upper and lower hemispheres’, and ‘maintain a sharp and straight edge’ must be completed in order to achieve the first sub-focus. Another set of hierarchically ordered sub-focus is constructed by considering how
upper and lower hemispheres are established. In this way, each hierarchical diagram branched into hierarchical instructions for lithic manufacture. Sub-foci which related to one another were joined with arrows to form phases. Continuing with the biface example, the sub-focus ‘create a pointed or ovate tip’ is joined to the sub-focus ‘create a suitable outline shape’ because the first must be achieved in order to complete the latter. Similarly, the sub-focus ‘remove protuberances’ is linked to the sub-focus, ‘thin the biface’. These two sets of linked sub-foci are not linked within the same phase, however, as they relate to different tasks (i.e. shaping and thinning respectively) which must be completed in order to achieve the primary focus. In the reduction sequences with no hierarchical complexity however, these sub-foci were organised two-dimensionally (i.e. with only one phase).

Additionally, while assigning tasks observed in the knapping footage to particular sub-foci may be a subjective task, this was done via consultation with the expert knappers. Should other analysts construct hierarchical diagrams from this knapping footage, slight variation may exist. However, regardless of this subjective interpretative component of the analysis, there are a number of immutable hierarchies involved in certain technologies. For example, Levallois technology invariably involves establishing a lower surface, an upper surface and a strong platform for final flake removal. Regardless of who performs the knapping and analysis, these three hierarchically distinct tasks must be present. To minimise any remaining subjectivity, knapping footage of two expert knappers were incorporated in this study and both knappers followed well established reduction sequences from archaeological, ethnographic and experimental sources.

6.2.4 Quantitative Comparisons
In addition to these visual comparisons of behavioural complexity provided by the hierarchical diagrams, this experiment provides quantitative assessments of the degree of hierarchical organisation. For example, the number of sub-foci in each diagram can be counted, an approach similar to that employed by Perreault et al. (2013). However, their quantification ignores the hierarchical structure of the sub-foci. A more detailed approach is to consider the hierarchical depth and breadth. As described in Figure 6.2, depth refers to the number of phases in a diagram, while breadth refers to the number of sub-foci in the longest phase. These variables reflect the number of phases and the maximum length of these phases on which a knapper must simultaneously concentrate. More complex technologies should involve more phases and longer phases.

These metrics only examine hierarchical complexity on a technology-wide scale. To further examine hierarchical depth and breadth, we can also observe how these variables change throughout each knapping sequence. When analysing the footage of the knappers, the enacted sub-focus was
recorded, along with its duration using the video player timestamp. Each time the knapper’s attention shifted to a new sub-focus, this new sub-focus was recorded, as was the time of this shift. When completing a particular sub-focus of a phase, a knapper must also be aware of all other downstream sub-foci within that phase (Wynn and Coolidge 2010b). Therefore, each downstream sub-focus was also recorded. A knapper can select any sub-focus in a phase, and therefore enact only a portion of phase, involving that sub-focus and the downstream sub-foci.

In terms of hierarchical depth, the number of phases comprise the mental palette from which a knapper may choose a course of action. Throughout a knapping sequence, a knapper enacts a certain phase, completes a portion of that phase and then chooses another phase. Repeating this sequence of cognitive choices results in the completion of the primary goal. More complex technologies should require more frequent changes to the active phase. The frequency of phase changes is therefore an important variable in reconstructing hierarchical complexity and is calculated here by recording the duration of each phase to the nearest tenth of a second.

Hierarchical breadth can similarly be reconstructed by examining the length of phases enacted throughout the sequence. By tallying the active sub-focus and all downstream sub-foci, the length of the phase can be calculated. As knapping proceeds, different phases of different lengths are selected until the primary goal is achieved. Technologies which possess longer phases require the knapper to hold more active sub-foci in their attention, and therefore involve more hierarchical complexity.

6.3 Results

6.3.1 Hierarchical Diagrams

Figures 6.3-7 show one hierarchical diagram from each of the five knapping strategies under investigation. Additional hierarchical diagrams which show further variation within technologies and include the diagrams produced from the footage of the second expert knapper are provided in the supplementary material. These diagrams show that bipolar (Figure 6.3) knapping involves a very short problem-solution distance, with only one phase and a narrow breadth of hierarchical complexity. As anticipated, bipolar knapping involves little conceptual distance from the overarching goal of detaching bipolar flakes. Discoidal knapping (Figure 6.4) requires a much greater breadth of hierarchical complexity (4 sub-foci) than bipolar.
Figure 6.3. Hierarchical diagram of the cognitive processes involved in bipolar knapping (Bipolar 1 and 2).

Figure 6.4. Hierarchical diagram of the cognitive processes involved in discoidal knapping (Discoidal 1, 2 and 3).

In comparison with discoidal technology, the hierarchical diagrams of biface manufacture (Figure 6.5) show an increase in both the breadth and depth of hierarchical organisation. Compared with Late Acheulean biface manufacture (Figure 6.5b), the Early Acheulean biface experiments (Figure 6.5a) involved fewer sub-foci and phases owing to the lack of platform preparation and modification to the plane of intersection. Refer to the supplementary material for the hierarchical diagram for the Abbevillian style biface with a ficron tip (Biface 2).
Figures 6.6 (a and b) demonstrate that the three variants of the Levallois method examined in this study are more hierarchically complex than any of the biface variants. Unlike bifacial knapping which involves two equivalent hemispheres, Levallois cores possess two hierarchically ordered hemispheres, each with a different function. This hierarchical ordering is represented in the diagrams by two of the three sub-foci which branch from the first sub-focus. Only once these hierarchically ordered tasks are completed can the platform on the lower hemisphere be prepared, further adding to the hierarchical complexity of Levallois knapping. This branching web of interacting sub-foci indicates that Levallois knapping involves many sub-foci and phases being indexed or stored for later use. Refer to the supplementary material for other examples of recurrent Levallois (Levallois 3) and two examples of recurrent point Levallois (Levallois 4 and 5).
Figure 6.6. Hierarchical diagrams of the cognitive processes in preferential Levallois knapping (a: Levallois 1) and recurrent preferential Levallois knapping (b: Levallois 2 and 6).
Finally, Figure 6.7 reveals that in comparison with Levallois knapping, blade core reduction involves fewer sub-foci and that these sub-foci are less interconnected. Therefore, prismatic blade core knapping requires less hierarchical depth and breadth compared with Levallois technology. Refer to the supplementary material for a hierarchical diagram of unidirectional crested blade knapping.

![Hierarchical diagrams of the cognitive processes involved in unidirectional blade core knapping (a: Blade 1 and 2) and bidirectional blade knapping (b: Blade 4).](image)

**Figure 6.7.** Hierarchical diagrams of the cognitive processes involved in unidirectional blade core knapping (a: Blade 1 and 2) and bidirectional blade knapping (b: Blade 4).

In order to test the role the idiosyncrasies of different knappers have on these results, hierarchical diagrams were similarly constructed for the four experiments conducted by the second expert knapper. Table 6.1 shows a summary of all the hierarchical diagrams constructed in this study and reveals little difference in the diagrams representing the two expert knappers. For example, the hierarchical diagrams for Discoidal 1 and 2 (Figure 6.4), Biface 3 and 4 (Figure 6.5b) and Levallois 2 (Figure 6.6b) were identical to the hierarchical diagrams constructed from the second expert’s footage. However, minor variations exist between the two knappers in the diagrams constructed for
the bidirectional crested blade iterations. The sequence conducted by the second knapper involved one additional sub-foci and one additional phase. Despite this variation, it appears that the idiosyncrasies of individual knappers has little bearing on the results for behavioural complexity. These continuities and differences can be explored further via the quantitative measures of behavioural complexity.

Table 6.1. Summary of the 21 reduction sequences conducted, including figure numbers and quantitative features of the hierarchical diagrams (hierarchical depth and breadth). The last three columns are a quantitative summary of the entire reduction sequence, including the total number of phases enacted throughout the sequence, the mean duration of those phases in seconds, and the mean number of sub-foci in those phases throughout the sequence (hierarchical breadth). Iterations in bold refer to those knapped by the second expert.

<table>
<thead>
<tr>
<th>Technology</th>
<th>Fig</th>
<th>Number of Sub-Foci in Diagram</th>
<th>Hierarchical Depth</th>
<th>Hierarchical Breadth</th>
<th># Phases Through Sequence</th>
<th>Mean Phase Time (sec)</th>
<th>Mean Hierarchical breadth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bipolar 1</td>
<td>6.3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>205.11</td>
<td>2</td>
</tr>
<tr>
<td>Bipolar 2</td>
<td>6.3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>111.39</td>
<td>2</td>
</tr>
<tr>
<td>Discoidal 1</td>
<td>6.4</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>103.92</td>
<td>2.5</td>
</tr>
<tr>
<td>Discoidal 2</td>
<td>6.4</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>184.16</td>
<td>2.5</td>
</tr>
<tr>
<td>Discoidal 3</td>
<td>6.4</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>61.48</td>
<td>2.5</td>
</tr>
<tr>
<td>Biface 1 (Early Acheulean)</td>
<td>6.5a</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>12</td>
<td>14.41</td>
<td>3.67</td>
</tr>
<tr>
<td>Biface 2 (Early Acheulean – ficon)</td>
<td>S1</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>10</td>
<td>38.66</td>
<td>3.3</td>
</tr>
<tr>
<td>Biface 3 (Late Acheulean)</td>
<td>6.5b</td>
<td>11</td>
<td>4</td>
<td>6</td>
<td>38</td>
<td>33.49</td>
<td>3.97</td>
</tr>
<tr>
<td>Biface 4 (Late Acheulean)</td>
<td>6.5b</td>
<td>11</td>
<td>4</td>
<td>6</td>
<td>36</td>
<td>22.37</td>
<td>3.94</td>
</tr>
<tr>
<td>Biface 5 (Late Acheulean)</td>
<td>6.5b</td>
<td>11</td>
<td>4</td>
<td>6</td>
<td>74</td>
<td>23.45</td>
<td>4.09</td>
</tr>
<tr>
<td>Levallois 1 (Preferential)</td>
<td>6.6a</td>
<td>18</td>
<td>7</td>
<td>6</td>
<td>62</td>
<td>9.08</td>
<td>4.53</td>
</tr>
<tr>
<td>Levallois 2 (Recurrent Preferential)</td>
<td>6.6b</td>
<td>19</td>
<td>7</td>
<td>6</td>
<td>133</td>
<td>12.33</td>
<td>4.65</td>
</tr>
<tr>
<td>Levallois 3 (Recurrent Preferential)</td>
<td>S2</td>
<td>16</td>
<td>6</td>
<td>6</td>
<td>99</td>
<td>12.45</td>
<td>4.64</td>
</tr>
<tr>
<td>Levallois 4 (Recurrent Point)</td>
<td>S3</td>
<td>24</td>
<td>10</td>
<td>6</td>
<td>54</td>
<td>14.76</td>
<td>4.5</td>
</tr>
<tr>
<td>Levallois 5 (Recurrent Point)</td>
<td>S4</td>
<td>23</td>
<td>10</td>
<td>6</td>
<td>47</td>
<td>12.91</td>
<td>4.28</td>
</tr>
<tr>
<td>Levallois 6 (Recurrent Preferential)</td>
<td>6.6b</td>
<td>19</td>
<td>7</td>
<td>6</td>
<td>58</td>
<td>13.05</td>
<td>4.84</td>
</tr>
<tr>
<td>Blade 1 (Unidirectional)</td>
<td>6.7a</td>
<td>10</td>
<td>4</td>
<td>5</td>
<td>31</td>
<td>20.12</td>
<td>3.13</td>
</tr>
<tr>
<td>Blade 2 (Unidirectional)</td>
<td>6.7a</td>
<td>10</td>
<td>4</td>
<td>5</td>
<td>28</td>
<td>30.46</td>
<td>2.75</td>
</tr>
<tr>
<td>Blade 3 (Unidirectional Crested)</td>
<td>S5</td>
<td>11</td>
<td>5</td>
<td>5</td>
<td>40</td>
<td>26.50</td>
<td>2.95</td>
</tr>
<tr>
<td>Blade 4 (Bidirectional Crested)</td>
<td>6.7b</td>
<td>11</td>
<td>5</td>
<td>5</td>
<td>50</td>
<td>34.36</td>
<td>3.16</td>
</tr>
<tr>
<td>Blade 5 (Bidirectional Crested)</td>
<td>S6</td>
<td>12</td>
<td>6</td>
<td>5</td>
<td>91</td>
<td>29.48</td>
<td>3.43</td>
</tr>
</tbody>
</table>

6.3.2 Quantitative Comparisons

The simplest comparison of the complexity of these hierarchical diagrams is the number of sub-foci involved, and is a similar method to that adopted by Perreault et al. (2013). Table 6.1 shows that Levallois knapping involves the most sub-foci, followed by blade, biface, discoidal and bipolar knapping. Similarly, the hierarchical organisation follows this same trend and is plotted in Figure 6.8. This chart represents both the hierarchical depth (number of phases in diagrams) and
hierarchical breadth (number of sub-foci in longest phase) of each knapping iteration. Therefore, the number of sub-foci and phases on which a knapper is concentrating and the number retained for later use is highest in Levallois knapping.

Figure 6.8. Clustered bar chart showing the hierarchical breadth and depth of each diagram. Asterisks denote repetitions conducted by the second expert knapper. Summary data for this chart can be found in Table 6.1.

The footage from these experiments were not only annotated into hierarchical diagrams, but were also used to extract detailed metrics about the number and nature of the features of the diagrams. For example, the frequency of phase changes, or the time spent on each phase, was calculated by recording the duration of each phase to the nearest tenth of a second. Technologies which involve a small problem-solution distance, and therefore little behavioural complexity, are expected to have a low frequency of phase changes. In other words, knappers should spend more time on particular phases for simpler technologies. On the other hand, it is anticipated that knappers must shift their attention among phases relatively frequently for more complex technologies. Figure 6.9 shows the result of plotting the time spent on each phase for each technology. Lower values indicate a higher frequency of phase changes.
Figure 6.9. Box-plots showing the frequency of phase changes for each reduction sequence, calculated by the time taken for each phase. A logarithmic (base 10) scale was used to accommodate the large values of the bipolar and discoidal iterations. Asterisks denote repetitions conducted by the second expert knapper. Summary data for this chart can be found in Table 6.1.

Bipolar knapping involves the least frequent changes to the active phase, but possesses too small a sample size to be statistically compared. When the other four technologies are considered together, a Kruskal-Wallis test (H = 115.7, d.f. = 868, p < 0.001) and post-hoc Mann-Whitney U-tests with Bonferroni corrections reveal that discoidal knapping involves the next least frequent phase changes compared to biface (U = 61, d.f. = 175, p = 0.0015), blade (U = 90, d.f. = 245, p = 0.0015) and Levallois knapping (U = 15, d.f. = 458, p < 0.001). Interestingly, the biface and blade iterations involve statistically equivalent rates of phase changes (U = 17820, d.f. = 409, p = 0.18). Once again, Levallois knapping involves the greatest complexity of technologies examined in this study, with the most frequent rate of phase changes compared with the combined values for biface (U = 27130, d.f. = 622, p < 0.001) and blade (U = 30050, d.f. = 692, p < 0.001) knapping.

In order to test whether the two knappers differ in terms of the frequency of phase changes while knapping, the average amount of time spent on phases by the two knappers was compared for each specific technology. The Late Acheulean bifaces of the first (Biface 3 and 4) and second expert (Biface 5) involved a comparable rate of phase changes (U = 2615.5, d.f. = 147, p = 0.64). So too did the crested bidirectional blade knapping of the first (Blade 4) and second (Blade 5) expert (U = 2180, d.f. = 140, p = 0.68). Likewise, no significant difference occurs between the first (Levallois 2
and 3) and second (Levallois 6) expert for recurrent preferential knapping \(U = 6315, \text{ d.f.} = 289, p = 0.47\). Thus far, the variation of individual knappers appears to have had no confounding influence on the results of this study.

Another means of quantifying the problem-solution distance is to examine the length of phases (number of active sub-foci) throughout the reduction sequence. As reduction proceeds, the knapper’s attention shifts from one phase to another, and these phases are of variable lengths. More hierarchically complex knapping sequences should possess longer phases. Technologies which necessitate longer phases require the knapper to hold more active sub-foci in their attention, and thereby require greater behavioural complexity. Figure 6.10 shows the mean length of the phases enacted in each knapping iteration. A clear upward trend is visible through the sequence of bipolar, discoidal, blade, biface and Levallois knapping. The only deviation in this trend compared to previous results, is that biface manufacture appears more complex than blade knapping.

![Figure 6.10](image.png)

**Figure 6.10.** Bar chart with one standard error bars showing the mean hierarchical breadth throughout each reduction sequence. The hierarchical breadth represents the number of active sub-foci in a phase. Asterisks denote repetitions conducted by the second expert knapper. Summary data for this chart can be found in Table 6.1.

When each technology is grouped together and compared to other technologies, the discoidal iterations involve significantly shorter phases compared to the biface \(\chi^2 = 88.80, \text{ d.f.} = 4, p < 0.001\) and Levallois \(\chi^2 = 32.36, \text{ d.f.} = 4, p < 0.001\) iterations, but not when compared to blade knapping \(\chi^2 = 3.37, \text{ d.f.} = 4, p = 0.50\). Meanwhile, blade knapping involves shorter phases.
compared with biface ($\chi^2 = 93.59$, d.f. = 5, p < 0.001) and Levallois knapping ($\chi^2 = 93.59$, d.f. = 5, p < 0.001). Finally, Levallois knapping again requires the greatest complexity, possessing significantly longer phases compared with biface technology ($\chi^2 = 99.53$, d.f. = 4, p < 0.001).

As with all previous metrics of hierarchical complexity, we wish to ascertain whether this trend is caused by actual technological differences or by individual knapper variation. When comparing the crested bidirectional blade iterations of both knappers (Blade 4 and 5), no significant difference occurs ($\chi^2 = 6.89$, d.f. = 4, p = 0.14). The same is true when comparing the recurrent preferential Levallois sequences of the first (Levallois 2 and 3) and second (Levallois 6) knapper ($\chi^2 = 4.51$, d.f. = 4, p = 0.34). Interestingly, the lengths of phases employed by the two knappers during biface manufacture (Biface 3 and 4 versus 5) were significantly different ($\chi^2 = 21.31$, d.f. = 3, p < 0.001). This may mean that the first knapper devised a less cognitively taxing approach to Late Acheulean biface manufacture. Alternatively, this slight difference may reflect raw material impurities that can necessitate additional problem solving to rectify.

According to the various metrics used to model hierarchical organisation, bipolar and discoidal technologies required the least and second least hierarchical complexity respectively. Biface manufacture displayed comparable or even higher levels of hierarchical complexity compared with unidirectional and bidirectional percussive blade knapping. Finally, all measures of the problem-solution distance indicate that Levallois knapping requires the greatest hierarchical organisation of any technology examined in this study.

6.4 Discussion

Hierarchical organisation has been used here as a proxy measure for behavioural complexity. It is therefore important that the method employed to quantify hierarchical complexity is not unduly influenced by confounding variables. The foremost potential source of confounding influence is the prospect for different knappers to achieve a comparable lithic end-product following different sequences of flake removals. It is therefore important that we focus only on metrics that reflect differences between technologies, rather than differences in the approaches of different knappers. In the crested bidirectional blade iterations for example, the second knapper enacted nearly twice the number of phases compared to the first knapper. The key results of this study however, frequency of phase change and phase length, were analogous and statistically indistinguishable between the two knappers (Table 6.1). It appears that the second knapper adopted a more iterative approach to reducing their blade core, employing more phases throughout the knapping sequence. Importantly, our results suggest that this variation in knapper style had no discernible influence on the
quantification of hierarchical organisation. The same can be said of all other technologies under examination here. Natural variation in approaches to stone tool making is to be expected. Choices relating to sequence of removals, size of removals, core maintenance and hammer selection vary considerably among different knappers, present and past. However, with more variation manifesting between different technologies than between the two knappers, we can conclude that the method employed in this paper discerns real differences in the hierarchical organisation involved in different lithic technologies.

The hierarchical diagrams and the quantitative analyses showed a consistent pattern of increasingly complex behaviour through the sequence of bipolar, discoidal, blade, biface and Levallois knapping. As anticipated, bipolar knapping involved very little behavioural complexity. Discoidal knapping required a greater number of sub-foci linked in series. This supports the idea that some Oldowan knappers possessed knapping schemas in which strikes were intended to both produce flakes and to maintain platforms with which to produce more flakes (de la Torre 2004).

Biface knapping was found to be significantly more complex than the discoidal cores which represented the pinnacle of Oldowan knapped technology. This resonates with previous evidence of the hierarchical and cognitive complexity involved in biface manufacture, as well as the suggestion that the origins of the Acheulean is an important cognitive threshold in human evolution (Gowlett 1986; Mahaney 2014; Stout et al. 2014; Stout and Chaminade 2012; Stout et al. 2011; Stout et al. 2008; Wynn 1995). In particular, it has been argued that greater fidelity in the transmission of technology allowed for more hierarchically complex sequences to be transmitted in the Acheulean (Shipton 2010; Shipton and Nielsen 2015). The biface sequences employing platform preparation and shifting the plane of intersection were more complex than those without these strategies, suggesting Late Acheulean knapping may be more cognitively demanding than that of the Early Acheulean.

According to all the measures employed in this analysis, the five Levallois iterations were found to be the most complex. Compared with the other strategies, Levallois knapping involved a greater number of sub-foci and phase changes, and increased hierarchical breadth and depth. The statistical tests demonstrated that Levallois technology involved the most rapid rate of phase changes and the highest hierarchical breadth throughout the sequences. This supports those that have emphasised the complexity of Levallois technology as well as the hypothesis that a major component of the Acheulean to Middle Palaeolithic transition was an increase in hierarchical behaviour and cognition.
Lastly, the three prismatic blade core iterations were found to be less behaviourally complex than the Levallois iterations. Therefore, this present study confirms the suspicions of Bar-Yosef and Kuhn (1999) that blade production does not signpost a leap in cognitive complexity.

It should be noted that this experiment only dealt with blank production and not retouch or hafting techniques, which may have been more complex in the Upper Palaeolithic. It would therefore be worthwhile examining the behavioural complexity involved in highly elaborate retouch technology and hafting. At the very least, Levallois technology represents the most cognitively demanding blank production technology possessed by hominins until the point of Neanderthal extinction. Hence we can argue that in terms of blank production, Neanderthals shared a similar level of behavioural complexity with that of their contemporaries, *Homo sapiens*.

Having examined the problem-solution distances of reduction sequences, it would be worthwhile investigating the extent the problem-solution distance would be protracted with the addition of other components that have been documented at various points in the archaeological record. Incorporating the processes of tool-stone procurement, heat treating, hafting, hunting, butchery and the manufacture of other tools would dramatically increase the problem-solution distances of the technologies investigated in this project. Sourcing of raw materials for hafting, processing these materials and combining these components into a single tool would greatly increase the number of sub-foci involved (Ambrose 2010; Carvalho et al. 2009; Haidle 2009; 2010; 2012; Lombard 2012; Lombard and Haidle 2012).

This study sought to reconstruct the behavioural complexity required for the production of different lithic technologies by estimating the extent of hierarchical organisation. It is also possible that evolving cognitive mechanisms are underlying this variation in hierarchical complexity. If so, then the trends of increasing hierarchical and behavioural complexity may also reflect increasing cognitive complexity.

The concept of ‘working memory’ has been flagged as an important cognitive mechanism involved in knapping (Ambrose 2010; Belfer-Cohen and Hovers 2010; Haidle 2010; Reuland 2010; Rossano 2010; Wadley 2010; Welshon 2010; Wynn and Coolidge 2010a). Working memory is the capacity to integrate moment-to-moment perception with archival information (Baddeley 1988; 1992; 2000;
The practice required to master stone knapping and the inherent unpredictability of individual knapping sequences (Bamforth and Finlay 2008; Bleed 2008; Geribás et al. 2010; Nonaka et al. 2010; Stout 2002) make knapping a good candidate for an activity that is demanding on working memory. Each flaking strategy in the mind of the knapper must be continually reappraised and adapted in light of the somewhat unpredictable nature of flake removals and the encounter of any internal imperfections in the stone.

In particular, the short-term aspect of working memory, the ‘episodic buffer’ (Baddeley 2000), is likely also involved. The episodic buffer acts as a storage system for short-term memory. It is a kind of palette, which facilitates the storage and access of multiple concepts simultaneously (Baddeley 2000; 2001; Coolidge and Wynn 2005; 2008; Haidle 2010; Wynn et al. 2009; Wynn and Coolidge 2010a). Information stored in the episodic buffer is thought to be consciously retrieved during an activity. Lithic technologies with longer phases and more frequent phase changes, such as Levallois knapping, likely require near constant retrieval and storage of information pertaining to the changing form of the core.

Aspects of longer-term memory, such as procedural memory are other structures likely required for hierarchically intensive knapping. Procedural memory is associated with unconsciously and automatically accessed skills and motor action (Ashby et al. 2010; Beilock et al. 2002; Doyon et al. 2009; Ullman 2001; 2004; 2016; Ullman et al. 1997). Procedural memories have been identified as a likely component of complex manual tasks such as knapping (Coolidge and Wynn 2008; Wynn et al. 2009), with the repetitive and iterative nature of knapping encoding these sequences of actions into memory.

Some have sought to synthesise and apply to the archaeological record both short- and long-term components of memory, such as the concept of expert cognition (Wynn and Coolidge 2004; 2010b). Others have identified key structures of the brain such as the cerebellum and prefrontal cortex which influence our capacity for short- and long-term memory (Asaad et al. 1998; Barton 2012; Gunz et al. 2012; Kane and Engle 2002; Stout and Chaminade 2007; 2009; 2012; Stout et al. 2011; Stout et al. 2008). This research makes clear that a holistic approach to cognition is needed. For example, Levallois knapping involved the greatest number of iterative sequences, as well as the greatest number and length of phases. These components of the knapping sequence directly relate to both the extent of procedural memory and working memory respectively. Evolutionary increases in procedural and working memory capacity may be key cognitive changes underlying the evolution of knapping technology from the Oldowan to the Middle Palaeolithic.
6.5 Conclusion

Palaeoanthropological attempts at charting the emergence of complex behaviour throughout our evolutionary history have typically used traces of symbolism, syntactic language or technology in the archaeological record. This study has offered a step forward in the latter approach by tracking the evolution of the hierarchical aspects of stone tool manufacture. Our results indicate increasing complexity between the Oldowan and the Early Acheulean, between the Early and Late Acheulean, between the Late Acheulean and the Middle Palaeolithic, but not between the Middle and Upper Palaeolithic.

Recent neuropsychological research has found hierarchical organisation to be a common component in both stone tool production and syntactic language (Greenfield 1991; Higuchi et al. 2009; Mahaney 2014; Morgan et al. 2015; Stout and Chaminade 2012; Stout et al. 2008; Uomini and Meyer 2013), with some arguing for a coevolution of complex language and tool making (Ambrose 2010; Morgan et al. 2015). The evidence outlined in this study for the gradual evolution of hierarchical complexity in stone tool manufacture may therefore also reflect the evolution of the hierarchical elements of syntactic language. As language is not highly archaeologically visible, the ubiquitous stone tool record may offer a means of charting the emergence of syntactic language. In this experiment, Levallois knapping required the greatest extent of hierarchical behaviour, providing evidence that important cognitive abilities required for complex language were likely present from the beginning of the Middle Palaeolithic or Middle Stone Age.
CHAPTER 7:
Conclusion

7.1 Addressing Research Questions

The previous four chapters outline four different studies that reveal a modicum of the diversity in knapping experiments. They are diverse in their archaeological applicability, as well as in the scale of their scope, methodological control and interpretations. As such, they offer an opportunity to explore the best-practice model of lithic experimentation outlined in Chapter 2 and to more closely examine how experimental knapping is best embedded within the broader archaeological research process. Doing so allows the research questions established in Chapter 1 to be addressed.

7.1.1 Research Question 1: Archaeological Applications of Experimental Knapping

The first research question asked ‘which aspects of the archaeological record are informed by knapping experiments?’ Accordingly, the archaeological utility and significance of each chapter is considered here.

The research outlined in Chapter 3 sought to provide a more reliable alternative to the typical method of flake platform area measurement. Instead of rectangular approximations of platforms, other geometric approximations were found to be significantly more reliable, and were even statistically indistinguishable from 3D scans of platform area. The null hypothesis outlined in Chapter 2, that the new method of platform measurement would be no more accurate or precise than the existing method, can therefore be wholly rejected. This new method of platform measurement can serve archaeologists in scenarios where 3D scanning is unfeasible and can contribute to reconstructing the reduction intensity of any tool type with intact platforms. Doing so yields a range of interpretations surrounding past human behaviour, including technological organisation, raw material consumption, mobility, subsistence, land-use and more.

As with the previous case-study, Chapter 4 concerns improving methods of measuring the reduction intensity of artefacts. Specifically, this study sought to develop and test a reduction intensity metric for the hitherto overlooked tool-type of backed blades. An experimental sample of blade blanks that were subsequently retouched into backed blades allowed the formulation of allometric relationships which in turn provided a reconstruction of original blade blank size. These allometric relationships represent a reliable measure of backed blade reduction intensity and were applied to an archaeological sample of backed blades from the Neolithic site of Boncuklu, Turkey. The
archaeological context at Boncuklu, dated to a transitional phase in which raw material was remarkably scarce, led to the hypothesis that the production of backed blades contributed to preserving their limited raw material and therefore assisted in the successful inhabitation of the site during transitions in their subsistence practices. The seemingly intentional optimisation and efficiency of blade blank consumption patterns lends support to this hypothesis. Additional confirmation of this hypothesis is provided by existing studies that identify backed blades as efficient, reliable and maintainable (Clarkson et al. in press; Hiscock 1994; 2002; Hiscock et al. 2011; Neeley 2002), as well as the subsequent chapter (Muller and Clarkson 2016a) which found small blades like those at Boncuklu to be remarkably efficient. In terms of this paper’s broader archaeological utility, this new reduction intensity metric could be applied to any blade-based backed artefact tradition, including those appearing at various times from the Middle Stone Age to the Neolithic in Europe, the Near East, Africa, South Asia, and parts of Australia.

Chapter 5 explored technological efficiency throughout human evolution by measuring the cutting edge efficiency of eight different technologies. This was accomplished using photogrammetry of the blanks removed from entire reduction sequences to provide the average length of sharp edge perimeter per gram of core for each technology. Based on previous experimentation (Eren et al. 2008; Jennings et al. 2010; Prasciunas 2007; Rasic and Andrefksy 2001), the null hypothesis was that no discernible differences in efficiency would exist among the eight technologies. However, with the results pointing to a generally ascending trend in efficiency over the course of lithic evolution, this null hypothesis can be rejected. Interestingly, the most significant increase in efficiency occurred at the transition from the Early Stone Age / Lower Palaeolithic to the Middle Stone Age / Middle Palaeolithic. The broader archaeological utility of this paper lies in the finding that no significant difference in cutting edge efficiency exists between the tool-kits of Neanderthals and their contemporaneous Homo sapiens.

Finally, as with the previous chapter, Chapter 6 investigates lithic technology on an evolutionary time-scale. Specifically, this chapter centres on the behavioural complexity required for the manufacture of five different technologies, spanning the Oldowan to the Upper Palaeolithic. Behavioural complexity is assessed by transcribing knapping footage into quantifications of the extent of hierarchical organisation involved in different technologies. As with Chapter 5, the null hypothesis is that no significant differences occur among the various technologies under investigation. This hypothesis was rejected based on the finding that significant differences occur among the five technologies, and that Levallois technology required the highest level of hierarchical organisation. It was therefore concluded that Neanderthals and their contemporaneous Homo
sapiens employed toolkits requiring equivalent levels of behavioural complexity in their manufacture. With hierarchical complexity involved in both tool-making and syntactic language (Greenfield 1991; Higuchi et al. 2009; Mahaney 2014; Morgan et al. 2015; Stout and Chaminade 2012; Stout et al. 2008; Uomini and Meyer 2013), the evidence for expanding levels of behavioural complexity required for tool manufacture may also reflect a growing capacity for complex language. With evidence of language preserving poorly in the archaeological record, the archaeological utility of this paper may go beyond mere technological interpretations.

In their comprehensive assessment of the utility of experimental knapping, Eren et al. (2016) identified three key areas most often aided by lithic experiments; method validation, hypothesis testing and predictive modelling. These four case-studies satisfy the first two of these research areas, with Chapters 3, 4 and 6 acting to validate new methods of analysis and Chapters 4, 5 and 6 serving to test hypotheses about the archaeological record. These case-studies have employed experimental knapping to further the research of flake measurements, reduction intensity metrics, raw material efficiency, and behavioural complexity respectively. However, single knapping experiments tend to influence more than one aspect of archaeological interpretation. The various ways in which these experiments have contributed to archaeological research are summarised in Table 7.1.

Table 7.1. Summary of the archaeological relevance of each case-study explored in this thesis.

<table>
<thead>
<tr>
<th>Utility</th>
<th>Case-Study</th>
<th>Chapter 3 (Platform Measurement)</th>
<th>Chapter 4 (Reduction Intensity)</th>
<th>Chapter 5 (Technological Efficiency)</th>
<th>Chapter 6 (Behavioural Complexity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method Validation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Lithic Analysis</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Efficiency</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduction Intensity</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evolutionary Trends</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

7.1.2 Research Question 2: How Experimental Knapping is Embedded within Archaeological Research

Table 7.1 not only represents a proportion of the wide range of possible applications advanced by experimental knapping, but shows the complex ways in which these studies are embedded in archaeological research. While Eren et al. (2016) identified hypothesis testing as a discrete research area of experimentation, it was stressed in Chapter 2 that hypothesis testing should be involved in any lithic experimentation. The generation of these hypotheses is reliant on a range of sources of evidence. As lithic experimentation should never stray far from archaeological relevance, perhaps
the foremost source of evidence with which to develop hypotheses is the archaeological record. Most commonly, this evidence takes the form of enigmatic patterns in lithic variability that are inexplicable with archaeological evidence alone. For example, the presence of overshot flakes on Clovis points (Eren et al. 2013) and the rate of errors for knappers of different skill levels (Eren et al. 2011b; Shelley 1990) led to influential experiments aimed at explaining these patterns of lithic variability. Other examples of external evidence include sources of analogy other than experimentation. For example, ethnoarchaeology has perhaps most notably underpinned the experiments of Shott (Shott and Sillitoe 2005; Shott and Weedman 2007) surrounding tool use-life and curation. Finally, refitting and the unchanging principles of physics and fracture mechanics that govern how stones fracture upon application of force also underpins hypothesis generation. Specifically, the study of fracture mechanics can establish experimental limits.

Although a key trend in knapping studies identified in Chapter 2 was the discipline’s shift from replication to experimentation, replicative studies can still play a role in experiments. By definition (established in Chapter 2), replicative knapping is inductive, and therefore not a component of the hypothetico-deductive model of science. While inductive reasoning is considered less powerful and reliable than deductive reasoning, instances of replicative knapping can serve to inform the intuitions of experimental knappers, and in turn help generate suitable hypotheses. While intuition alone is insufficient to interpret experimental or archaeological evidence (Eren et al. 2016; Shea 2011), its role should not be dismissed entirely. Any experiment involving an expert knapper can never be wholly divorced from their decades of experience and intuition, and instead this experience can be directed towards suitable hypothesis development.

Figure 7.1 synthesises these various sources of evidence that contribute to hypothesis generation in knapping experiments, and shows the complex ways these lines of evidence interact. This flow-chart is centred on experimentation, and therefore shows the role of knapping experiments on archaeological research, and archaeological research’s role on experimentation. This iterative research flow, or feed-back loop of hypothesis development and testing, also includes the three research areas identified by Eren et al. (2016). Hypothesis testing is clearly represented in the main flow of ideas, but the application of experiments to method validation and predictive models are also shown. The validation of experimental methods serves to strengthen this main interpretative sequence. While there is a feed-back loop between interpretations and hypothesis generation, a second feed-back loop involves the use of experimentation to create predictive models that can be applied to the archaeological record, which in turn inform the generation of new experimental hypotheses. While none of the case-studies presented in this thesis developed predictive models,
they did involve method validation and hypothesis testing. Although this flow-chart is presented as a closed system, if the entirety of archaeological and ethnoarchaeological research was included, the chart would expand considerably. Figure 7.1 is intended to simply show some of the ways experimentation is embedded within broader avenues of archaeological research.

**Figure 7.1.** A synthesis of the range of sources of evidence that contribute to hypothesis generation in knapping experiments and how this experimentation is embedded in a feedback loop of hypothesis generation, experimentation and hypothesis testing.

7.1.3 Research Question 3: Best-Practice Model of Knapping Experimentation

The primary goal of this thesis was to identify features of knapping experiments that contribute to an experiment’s validity. It was argued in Chapter 2 that the scale of interpretations borne from knapping experiments should be directly proportional and inversely proportional to the scale of scope and methodological control respectively. Plotting the theoretical relationships between scope, control and interpretations resulted in a two-dimensional plane that places theoretical limits on the kinds of interpretations that can be made depending on the initial experimental parameters of scope and control.

The four case-studies explored in this thesis were conducted with a diverse set of initial parameters, ranging from the specific and strictly controlled, to the broad and more loosely controlled. These
four case-studies were purposefully ordered in this thesis according to their ascension on this two-dimensional plane. Situated at its lower end are studies with a narrow initial scope and strict methodological control. These studies result in narrow and specific interpretations that will bear very few exceptions. Meanwhile, the upper extremity of the plane represents studies with broad scope and low control, resulting in broad, encompassing interpretations. However, as the breadth of interpretations increases, so too does the number of possible exceptions that will not conform to such broad generalisations. As was argued in Chapter 2, studies on this plane may differ markedly in their initial set-up and subsequent interpretations, but while they are situated near this plane, they possess equivalent validity. As a theoretical exercise, the approximately surmised location of each case-study has been plotted on this two-dimensional plane (Figure 7.2). These plots have been made approximately rather than empirically, and only offer a general gauge of the diversity of the four-case studies in terms of their experimental parameters and how well their interpretations conform to the best-practice model of experimental knapping.

Figure 7.2. The best-practice model for experimental knapping developed in Chapter 2 (Figure 2.1) with each case-study (Ch3-6) overlain, approximately representing where each experiment is situated relative to the two-dimensional plane.
Chapter 3 involved the precise measurement of a single flake attribute and therefore constituted a very narrowly defined scope. In terms of methodological control, while these flakes were made with free-hand knapping rather than a mechanical apparatus, the flakes were knapped by a single expert knapper unaware of the goals of the study. Additionally, the calliper measurements of platform area were conducted in double-blind conditions, where neither the authors nor participants of the inter-user variability portion of the experiment were aware of the results of the 3D scanned measurements of platform area. Therefore, Chapter 3 was conducted with a relatively high degree of experimental control. Coupled with the narrow scope of research, the interpretations should be correspondingly narrow. With the conclusions borne from Chapter 3 culminating in a verification of a new method of measurement, it can be argued that the interpretations were suitably narrow.

Chapter 4 contained an experiment that not only developed a new method of analysis but also applied this method to the archaeological record, thereby somewhat increasing the scope of analysis and loosening the extent of experimental control. The type of analysis, relatively simple calliper measurements and attribute recordings, maintains a relatively high level of methodological control. In contrast, a less well tested set of methods would involve less methodological control. The resultant conclusions involved interpretations about the nature of the archaeological record in tandem with other lines of archaeological evidence such as faunal and botanical research. As such, these interpretations were considerably broader than those made in Chapter 3.

Expanding the initial scope further, Chapter 5 involves estimating the efficiency of lithic technologies that span the Oldowan to the Neolithic, and therefore tracks technological change on an evolutionary time-scale. Meanwhile, the use of photogrammetry on a large sample of blanks affords a moderate level of methodological control. However, the inclusion of two knappers of differing skill level may increase the strength of the interpretations, but it slightly lessens the extent of methodological control. Combining an evolutionary scope with moderately well controlled methods results in medium-to-broad interpretations. With the results of this experiment informing the technological efficiency of tool-kits possessed by different hominin species, the scale of interpretations likely conforms to these parameters.

Lastly, like the previous case-study, Chapter 6 occurs on an evolutionary scale and therefore has a very broad temporal scope that spans the Oldowan to the Upper Palaeolithic. However, this scope is marginally lower than in Chapter 5, as the previous study included technologies from the Mesolithic and Neolithic. Meanwhile, the methodological control of this experiment is relatively low, as concepts like behaviour and cognition are difficult to precisely and accurately quantify compared
with technological efficiency, reduction intensity and flake attributes. However, several methodological choices were made to mitigate the influence of confounding factors and somewhat tighten the methodological control. For example, the inclusion of a second expert knapper was designed to test the role of idiosyncratic behaviour between individual knappers. Additionally, strict constraints were imposed on both knappers while performing the different reduction sequences. Despite these methodological constraints, the imprecise nature of cognition in the archaeological record means that the experimental control can be tightened only so far. With concepts like behaviour and cognition remaining enigmatic pursuits in archaeology, the breadth of interpretations is difficult to gauge. Many remain sceptical of archaeologists’ ability to infer cognition in the archaeological record. This study was intended to provide a novel approach at quantifying concepts like hierarchies, behaviour and cognition. While these quantifications were based on neurological and psychological research, any study of cognition in the past will be fraught with potential confounding factors. Therefore, taking a conservatively sceptical view of this study, this experiment has been plotted slightly above the two-dimensional plane in Figure 7.2. However, if further experimentation and archaeological research corroborates the findings of this experiment, then a less sceptical view of this study can be adopted, thereby situating the interpretations of the experiment nearer the two-dimensional plane.

As was posited in Chapter 2, different knapping experiments may differ markedly in their scope and methodological control, but they do not necessarily vary in their validity. Based on the generalisation that scope and control should be directly and indirectly proportional to the scale of possible interpretations respectively, studies that exist within these bounds theoretically possess equivalent validities. The spread of the four case-studies in Figure 7.2 shows a portion of the variability in initial parameters that experiments can possess and also supports the flexibility of this model. A sceptical view of the experiment in Chapter 6 situates this study slightly above the plane of Figure 7.2, meaning that there may be a discrepancy between the broad interpretations and the initial parameters. Future work could further refine the interpretations of this study, with experiments of narrower scope or tighter methodological control possibly reinforcing or disputing these interpretations.

7.2 Avenues for Future Research
As with most archaeological research, experimental knapping studies tend to reveal more questions than they answer, and these four case-studies are no exception. For example, the analysis of backed blades at Boneuklu revealed the possibility of an intentional efficient use of blade blanks within the backed blade reduction schema. Ongoing research has revealed the efficiencies and optimality of
other components of the tool-kit at Boncuklu. For example, data for the size of artefacts, reduction intensity, core exploitation, core exhaustion and bipolar knapping strategies all suggest an extreme degree of raw material rationing occurring at the site. This evidence is starkly contrasted with the presence of larger and more formalised lithic specimens in burial and cache contexts. Future studies will explore how the lithic raw material at Boncuklu played significant roles in different spheres of life at the site, from the more functional use of efficient technologies, to the less functional interment and caching behaviours. Additionally, experimentation is already underway to identify traces of high-velocity projectile use at Boncuklu. Experiments are being conducted to differentiate fracture patterns on small obsidian pieces from a range of functional and taphonomic processes. The highly fragmented assemblage at Boncuklu poses a challenge to such analyses, making the experimental differentiation of impact fractures, bipolar knapping, pièce esquillées use, and trampling particularly useful.

The experiments in Chapters 5 and 6 explored trends in lithic technology over the course of human evolution, namely efficiency and behaviour. Another key trend that likely influenced lithic variation on an evolutionary scale was the skill involved in the manufacture of different technologies. As was discussed in Chapter 2, much work has focussed on questions of skill in the archaeological record (Eren et al. 2011b; Eren et al. 2011c; Muller and Clarkson 2016a; Nichols and Allstadt 1978; Nonaka et al. 2010; Shelley 1990; Stout et al. 2014; Stout and Semaw 2006). However, little has been done to quantify and compare the relative amounts of skill required for the production of different technologies throughout human evolution. A method similar to that outlined in Chapter 6 could reveal some key trends in skill requirements over the last 3.3 million years. However, disentangling experimental evidence for relative levels of skill and cognition must be a primary task of such a study.

In terms of the discipline of experimental knapping more generally, some key future directions of research can be identified. For instance, it was discussed in Chapter 2 that both free-hand and mechanised knapping experiments can play complementary roles in hypothesis creation and testing. For example, knapping with a mechanised apparatus revealed key variables that influence flake size, such as platform attributes, billet type and velocity (Dibble 1997; Dibble and Pelcin 1995; Dibble and Rezek 2009; Dibble and Whittaker 1981; Lin et al. 2013; Magnani et al. 2014). These observations were later confirmed, and therefore strengthened, by more archaeologically realistic free-hand knapping (Bradbury et al. 2008; Clarkson and Hiscock 2011; Davis and Shea 1998; Muller and Clarkson 2014; Shott et al. 2000). This interplay between mechanised and free-hand experimentation offers the most profitable avenue for future experimental study, with the cycle of
hypothesis generation and testing being considerably bolstered if this cycle includes both styles of experiment and is linked to the archaeological record. Additionally, future work will likely involve a marginally more realistic mechanised style of knapping, such as a pivoting mechanical arm that more closely resembles the motion of free-hand knapping.

Lastly, the ever growing field of computer science offers burgeoning applications to both archaeology and experimental knapping. As was discussed in Chapter 2, the use of 2D and 3D photogrammetry and geometric morphometrics have increasingly played a role in lithic experimentation (Archer and Braun 2010; Clarkson 2013; Clarkson and Hiscock 2011; Clarkson et al. 2014; Grosman 2016; Grosman et al. 2011a; Grosman et al. 2014; Grosman et al. 2011b; Grosman et al. 2008; Li et al. 2015; 2016; Lin et al. 2010; Muller and Clarkson 2014; 2016b; Richardson et al. 2014; Shipton 2016; Shipton and Clarkson 2015a; 2015b; Sholts et al. 2012; Shott and Trail 2010; Zaidner and Grosman 2015). Undoubtedly, these analyses will only grow more complex with the introduction of more refined 3D methods and the application of machine-learning algorithms to archaeology.

7.3 Summary
This thesis has sought to demonstrate the archaeological utility of experimental knapping, reveal the complex flow of research within which experimentation is embedded, and above all, develop a best-practice approach to experimental knapping. It has been argued that the validity of lithic experimentation partly relies on the use of explicit, testable and falsifiable hypotheses. The hypotheses contained within the four case-studies above served to demonstrate the interpretive power of this deductive approach in experimental knapping. Foremost, however, this thesis has stressed the significance of marrying the scale of the scope, methodological control and subsequent interpretations. The number of experiments in lithic studies is already growing. If future experiments conform to this model, the number and efficacy of knapping experiments will likely only improve. In sum, adherence to sound experimental parameters and strict hypothesis testing, coupled with emerging techniques of lithic analysis, can facilitate powerful and robust archaeological interpretations.
References


Akerman, K. 2007 To make a point: ethnographic reality and the ethnographic and experimental replication of Australian macroblades known as Leilira. *Australian Archaeology* 64:23-34.


Ataman, K.A. 1989 The chipped stone assemblage of Can Hasan III: a study in typology, technology and function, Ph.D. Dissertation, University College London.


Bourguignon, L. 1997 Le Moustérien de type Quina: nouvelle définition d'une entité technique, Ph.D. Dissertation, Université Paris X-Nanterre.


Brindley, J. and C. Clarkson 2015 Beyond a suggestive morphology: were Wardaman stone points exclusively spear armatures? *Australian Archaeology* 81:12-23.


Callahan, E. 1987 *An evaluation of the lithic technology in Middle Sweden during the Mesolithic and Neolithic*. Sweden: Societas Archaeologica Upsaliensis.


Crabtree, D.E. 1970 Flaking stone with wooden implements: flaked stone artifacts from Palliaike, Chile, suggest that wooden flaking tools were used in the New World. *Science* 169(3941):146-153.


Cushing, F.H. 1879 Curious discoveries in regard to the manner of making flint implements by the aborigines and prehistoric inhabitants of America. *Engineering and Mining Journal* 28:91.


Heighway, Z. 2011 At the Cutting Edge of Core Technology: Modern Notions of Reduction Efficiency, BA(Hons), University of Queensland


Hiscock, P. and C. Clarkson 2007 Retouched notches at Combe Grenal (France) and the reduction hypothesis. *American Antiquity* 72(1):176-190.


Lombard, M. 2012 Thinking through the Middle Stone Age of sub-Saharan Africa. *Quaternary International* 270:140-155.


Low, M.A. 2015 Investigating standardisation in the form of backed artefacts at two sites in the Hunter River valley, NSW, Australia. *Australian Archaeology* 80:60-69.


Mackay, A. 2011 Potentially stylistic differences between backed artefacts from two nearby sites occupied 60,000 years before present in South Africa. *Journal of Anthropological Archaeology* 30:235-245.


Marks, A.E. and V.P. Chabai 2006 Stasis and change during the Crimean Middle Paleolithic. In E. Hovers and S.L. Kuhn (eds), *Transitions before the Transition: Evolution and Stability in the Middle Paleolithic and Middle Stone Age*, pp.121-135. New York: Springer.


Roe, B.E. and D.R. Just 2009 Internal and external validity in economics research: tradeoffs between experiments, field experiments, natural experiments, and field data. *American Journal of Agricultural Economics* 91(5):1266-1271.


Schoville, B.J. 2014 Testing a taphonomic predictive model of edge damage formation with Middle Stone Age points from Pinnacle Point Cave 13B and Die Kelders Cave 1, South Africa. *Journal of Archaeological Science* 48:84-95.


Semaw, S. 2000 The world’s oldest stone artefacts from Gona, Ethiopia: their implications for understanding stone technology and patterns of human evolution between 2·6–1·5 million years ago. *Journal of Archaeological Science* 27(12):1197-1214.


Wurz, S. and M. Lombard 2007 70 000-year-old geometric backed tools from the Howiesons Poort at Klasies River, South Africa: were they used for hunting? *Southern African Humanities* 19:1-16.


Zaidner, Y. and L. Grosman 2015 Middle Paleolithic sidescrapers were resharped or recycled? A view from Nesher Ramla, Israel. *Quaternary International* 361:178-187.