10. FAILURE AS TRUTH: AN AUTOPSY OF CRABTREE’S FOLSOM EXPERIMENTS

John E. Clark

It would appear that a law of mechanics would forbid the fluting of a Folsom projectile point. It would be easier to explain why the Folsom cannot be fluted rather than to describe how it was made. (Don Crabtree 1966:21)

Replication of flaked stone tool reduction technologies within strict scientific and experimental guidelines will be the only demonstrable method of understanding prehistoric behavior reflected by flintknapping. (J. Jeffrey Flenniken 1984:200)

Consider the hubris of scientific knapping’s undergirding assumption: modern experiments can reveal ancient truths. There is no stronger philosophical claim in archaeology for recovering the past. If parts of the past are knowable, as knappers claim, how is this possible? In the specific example of concern here, how can one discover how ancient Folsom points were made—or not made? And, how can one be sure that her or his discovery in the present was true of a past? I consider these dual epistemological challenges through a critical appraisal of Don Crabtree’s (1966) Folsom experiments, the foundational statement that established ground rules for experimental replication. Crabtree almost single-handedly moved lithic studies from taxonomy and description to explanatory concerns dealing with the manufacture of specific artifact forms, but as patron saint of replication studies he is generally esteemed so highly that his work receives more reverence than careful readership. Rather than exalt Crabtree’s speculations on knapping issues, I highlight their deficiencies. As some recent experimental work shows, most discoveries Crabtree claimed for the manufacture of Folsom points and related artifacts are erroneous or dubious (see Clark 1982; Flenniken 1978; Gryba 1988b, 1989; Sollberger 1985; Titmus and Clark 2002; and this volume). If accurate, Crabtree’s fabulous failure rate raises serious concerns about the soundness of his method. In addressing these questions, I take Crabtree’s experiments with Folsom points as indicative of his more general approach.

Some may find my objective inherently offensive because it seems to come at Crabtree’s expense. Offense taking at criticism, however, would only devalue Crabtree’s contribution and his personal quest for truth. Crabtree was never satisfied with his
own level of understanding of various technologies, and neither should we. I respect the Crabtree approach and wish to propel it towards greater rigor. It is worth mentioning that the information undercutting his claims derives from experiments similar to his. So all parties agree on general method and only dispute proposed facts. In knapping, as in science, errors become fatal only if left unrecognized and unattended. I believe more can be learned from Crabtree’s interpretive excesses than from his modest successes, and I pursue critique with this hope.

After first summarizing Crabtree’s Folsom experiments and major truth claims derived from them, I attempt to recover the implicit logic behind his experimental method and inferences. Crabtree got most things wrong, for reasons both within and beyond his control. But to profit from his miscues we must identify them and understand how and why they may have occurred. Such exercises are inherently normative because identification of mistakes automatically marks paths to prevention and resolution.

THE EXPERIMENTS

Mining the bedrock of Crabtree’s method requires examining details of his discoveries. Crabtree tested 11 general classes of knapping techniques for manufacturing Folsom points, with some sub-variants. All told, he provided information on 17 classes of experiments, but actual experiments numbered in the hundreds as many ideas had to be tested many times. Information provided for each experiment is uneven, with the more successful experiments meriting the greatest expository attention.

Details of Crabtree’s experimental approach are implicit but sufficiently clear in his article. Everything begins with interest in a question, in this instance, the question of how Folsom points were made. For Crabtree (1966: 6–9), following steps were to examine and describe the artifact(s) in question, hypothesize possible manufacturing techniques, conduct experiments with these techniques, compare the results of his experiments to the archaeological specimen(s), and then reject or accept experimental results based on their fit with archaeological parameters. Of the 17 procedures evaluated, Crabtree (1966) rejected all but three. He successfully fluted points by direct pressure with a chest crutch, clamp, and anvil rest and by indirect percussion with an anvil and clamp. A combination of the two might also serve. He rejected direct freehand percussion (with all kinds of percussors), direct percussion using an anvil or holding device, freehand indirect percussion (holding the preform in various ways), freehand pressure with long and short tools, and pressure with a chest crutch and clamp without an anvil rest.

The experiments cover an impressive range of techniques that match those described by William Henry Holmes (1919) for aboriginal America and first tested by H. Holmes Ellis (1940) in the Ohio Lithics Lab, where Crabtree received his formative knapping experience. Technical minutia listed by Crabtree are impressive, as are the staggering number of truth claims derived from his experiments. Unfortunately, his claims were precipitous. Subsequent experiments conducted by others, and following similar logic, demonstrate that Crabtree’s experimental points were not true replicas in terms of basic measurements of attributes (see Sollberger 1985), and many of the techniques he rejected as impossible have since been used to flute preforms. Eugene Gryba (1988b, 1989) successfully flutes long points with direct hand held pressure using a short antler tool; Gene Titmus (Chapter 12) and Bob Patten (Chapter 15) flute points with direct percussion with an antler baton; Kenneth Rozen (Chapter 14) flutes points with a simple holding device and a chest crutch; Phil Geib and Stan Ahler (Chapter 13) use indirect percussion for fluting; and Jeb Taylor is able to
flint points with indirect percussion while immobilizing the preform with his tied feet. Other variants exist. Most of these methods were demonstrated at the Folsom Workshop Conferences. They represent counterclaims to Crabtree's about the feasibility of various methods.

Viewed generically, it appears that almost any conceivable method for securing a preform and detaching a channel flake serves in fluting a Folsom, but such a view cannot withstand close scrutiny. The critical questions concern the criteria for identifying true replicas and justifiable arguments for making such identifications (see below). There is more to Folsom points than a flute down the belly of a short biface. Most claims for replication cannot be sustained because the duplicated forms deviate from the attributes of archaeological specimens. Crabtree's fluted points, and those of other replicators, fail to make the grade.

Crabtree advanced propositions about Folsom points concerning requisite skill, the manufacturing sequence, and the manufacturing technique. He considered the Folsom point "to be one of the most beautiful, practical, highly specialized, and, admittedly, one of the most difficult points to replicate. ... the making of this point probably took more time, patience and skill than any other projectile point of comparable size" (Crabtree 1966: 3). Skill levels involved both the fluting and marginal retouch of finished points.

The retouching on the margins of the Lindenmeier Folsom is equal in skill to the channel flaking ... These parallel marginal retouch flakes are as close as one twenty-sixth of an inch in width. ... Infinite skill is required to remove each of these diminutive micro-flakes, for each flake removal requires the same platform preparation, the same spacing, the same downward and outward pressure and the force must be applied each time at exactly the same angle. (Crabtree 1966: 5)

Crabtree's description of the manufacturing sequence for Folsom points is a bit difficult to disentangle because he often failed to distinguish between observations of experiments and assertions about ancient techniques. He imagined a reduction sequence that progressed from blanks to preforms by direct freehand percussion, both with a hammerstone and antler billet. After roughing out the preform it was carefully pressure flaked on both faces.

The first pressure flaking is not an attempt to produce uniformity, but is merely to remove any irregularities or step-fractures left by the percussion work. The preform must then be retouched again to make regular, uniform flakes over the entire surface of the artifact. This provides the smoothness and regularity necessary for removal of the channel flake. (ibid.: 18)

Using a tool with a very fine point, the edges are then pressure retouched by removal of a series of narrow, minute, parallel flakes. This results in an edge that is thick but very sharp, which serves a dual purpose. First, it will withstand the pressure of a holding device; and second, it gives strength to the projectile when it is finally completed. The edge at the basal portion of the artifact is ground smooth for additional strength ... The distal end of the artifact should be left rather blunt and almost as thick as the mid-section to provide for the beveling and polishing of the tip and still have enough strength to support the force of removing the fluting flake. It is this part of the point that will rest on the antler anvil during the fluting process and it must withstand the force necessary to remove the flute. (ibid.: 19)

Following stages relate to the fluting process and difficulties of creating isolated platforms at the concave base for removing a flute from each face. Crabtree provides detailed and careful instructions about this stage of the process.
The next stage in flute removal is the preparation of the first platform ... the base of the preform has been left either square or with a convexity and the worker must now isolate the platform from the tangs. ... The center portion of the base will be used for seating the pressure tool when the platform is completed. The platform must now be freed by applying the pressure tool on the opposite side of the base to remove, by a series of graduating pressure flakes, the material between the lateral edges and the area on which the fluting tool will rest (platform). A series of graduating flakes is removed by pressure from the side to be fluted, starting from the proposed tangs with the last, and longest, flake terminating at the median line to form a spine directly in line with the tip. The same procedure is repeated on the opposite side. The platform should then be left projecting slightly less than a quarter of an inch above the two concavities between the tangs and the platform. The projection must be freed on the side opposite the face to be fluted. ... The top of the platform is then polished to prevent crushing from the application of force. (ibid.)

The preform is then clamped at the front end between the two strips of wood and is positioned about 10 degrees from vertical ... The distal end of the polished, beveled tip will rest on the leading edge of the anvil. The anvil may be of any resilient material, but one must not use any unyielding substance. ... The polished distal end of the artifact must be held firmly against the anvil by means of the downward pressure from the tightened vise or clamp. (ibid.: 20)

To hold the vise stationary, the flaker must now stand on the clamp, with the chest crutch in place and the worker in a bending position. ... The shaft of the pressure tool must be vertically and directly in line with the median line of the artifact. ... Both hands are then placed on the shaft of the crutch at a position just opposite the knees. The knees may then assist the hands in controlling the outward pressure. Outward pressure is then gradually increased by the weight of the body and pressure from the knees until the platform parts from the base and the channel flake is pressed off to the tip of the projectile point. (ibid.: 20–21)

...the first channel flake has been removed in a satisfactory manner resulting in a flake scar on the artifact having the same character as that of a Lindenmeier Folsom. The half-fluted point is then removed from the clamp and a second platform is prepared on the opposite side in the same manner as the first. This second platform will, however, be slightly lower than the first. ... After the second platform is prepared and the tip reconstructed and polished, the half-fluted point is then placed in the clamp for the removal of the second flake. ... (ibid.: 21)

Crabtree was somewhat dissatisfied with his results of his experiments because they failed to supply sufficient evidence to winnow the candidates for the manufacturing sequence down to a single technique.

I am left with the disquieting fact that I can replicate the Lindenmeier Folsom by the use of two techniques and the nagging thought that, at this time, I cannot discard either method. Yet it is unlikely that this point was made by the use of two different techniques. My experiments indicate that this projectile point was made by either the indirect percussion with rest method, or the pressure with clamp and anvil technique. (ibid.: 22)

The preceding, salient claims for fabricating Folsom points presuppose many others. The final
and ignoring early steps he failed to understand Folsom fluting or retouch.

Crabtree (1966: 3) relied on one epoxy cast from the Denver series (Point B22/83) to guide his experiments and assess their output (see Chapter 12). It is a lovely point, but a final form. Another clear lesson apparent to all at the first Folsom Workshop Conference was the resharpening cycle of Folsom points (see Chapter 20). Beginning forms were about twice as long as final forms. The specimen Crabtree chose for his replication target was a “slug” on its last legs, a final form (see Chapter 13: Fig. 13.4a). This is to say that Crabtree did not understand final forms either. Consequently, he overidealized and complicated the replication task, striving for a level of perfection and difficulty not represented in the archaeological record. His need for complicated holding devices for removing channel flakes with chest pressure or indirect percussion related to his quest for perfection. This was a mistake he repeated with Mesoamerican obsidian blades a few years later (Crabtree 1968), and for the same reasons of starting with beautiful final forms viewed in private collections in the absence of any preliminary or intermediate forms for examination (Titmus and Clark 2002).

Crabtree tried to duplicate the ultimate aesthetic forms, and when he saw the clunkers and rejects from early stages of manufacture he did not (could not?) associate them with the finished product. Thus, in his Folsom experiments, he expended prodigious effort in trying to get perfectly-flaked preforms with parallel microflaking before fluting, never realizing that the fine retouch followed fluting and sometimes multiple resharpenings. Breaking such labor-intensive preforms in fluting merely increased the heartache of failed experiments. His mistakes in analysis and conception of the replication target are all resolvable issues. The main purpose of the Folsom Workshop Conferences was to bring analysts, expert knappers, and collections
together so issues of access and perception could be resolved through public dialog, and more directed replication experiments could ensue.

Beyond the noted lapses in preliminary analysis of archaeological collections, Crabtree’s (1966) foundational article promotes a family of errors of greater interest. These concern foundational logics for replication experiments themselves. How can one claim ancient knowledge through modern experiments? Crabtree made both weak and strong claims that various techniques were or were not used in the past. As apparent in counterclaims discussed in this volume, Crabtree dismissed failed techniques too hastily. Replication experiments necessarily foster an asymmetry in truth claims. A knapper is clearly justified in claiming that a given technique is one possible way ancient points might have been made once he or she has actually made legitimate replicas with the technique. In contrast, claims about knapping impossibilities based on failed experiments are not justified—at least not right away. There are so many ways and reasons an experiment may fail that strong inferences from failures require many more trials than do successes. Crabtree’s claim in the epigraph has things backwards. Confirmation is immediately clear, but disconfirmation requires protracted study. In Crabtree’s Folsom experiments, failures include those with the least number of trials, the reverse of logical necessity. Moreover, the various claims for successful and unsuccessful methods show clearly that Crabtree’s logical claims are all ad hominem, based on his own personal authority. A claim that a given technique could not work was Crabtree’s way of saying that he did not or could not get it to work. Given his skills and experience, his claims still carry considerable weight, but they do not constitute logical arguments.

Crabtree’s rehearsal of his Folsom experiences reaffirmed scientific knapping as a way to understand aspects of the past. Paradoxically, none of his substantive inferences has withstood the test of time, but his method has. His reconstruction of Folsom technology has been hard to reconcile with Folsom lifeways and the organization of technology because the proposed manufacturing procedure coupled the need for supreme manufacturing skill with extreme risk at the very final step in making these points (see Ingbar and Hofman 1999; Chapter 13), neither a practical nor smart tactic. If manufacture was so precarious, why would pragmatic herd hunters risk fluting their points, especially when away from sources of toolstone? Ongoing experiments are showing that things were not as extreme as Crabtree described. As contributors to this volume demonstrate, fluting requires a minimal toolkit and can be done consistently at moderate risk. Although the precise method, or methods, used to make Folsom points at any given time or place remain(s) to be demonstrated, the most likely possibilities are known. On the conceptual side, however, it is still not clear how anyone can justifiably claim that ancient points were made in a particular way.

LOGIC OF THE ARGUMENTS

I lack the training to be comfortable in the task of evaluating thoroughly points of logic and epistemology for Crabtree’s experimental approach. My more limited objectives here are to raise fundamental questions and initiate discussion on these important matters with the hope that others more qualified can complete the chore. For now it suffices to consider two of Crabtree’s truth claims and take them as representative of the types of claims asserted by scientific knappers.

Claim 1: “To my knowledge, no present-day flintknapper has ever really mastered the Folsom techniques, but my experiments have helped eliminate, for me, some of the methods purportedly used.” (Crabtree 1966: 9)

Claim 2: “I have tried every conceivable method of producing this fluted artifact and have, finally,
accepted two methods and find that a third technique has merit but needs further experimentation. Accepted methods are (1) fluting by direct pressure with rest; (2) fluting by indirect percussion with anvil and clamp; (3) combination of both.” (ibid.)

Crabtree’s basic method for discovering ancient manufacturing techniques was the process of elimination, or basic trial and error. Experiments with unacceptable outcomes led to a presumption that the technique involved was not used in the past to flute Folsom points. But to gain interpretive leverage from failures, one must understand why they occurred, a rather difficult task. Successful experiments identified those techniques most likely used. In the details of each experiment, Crabtree (1966) lists reasons why he was dissatisfied with the results of some, why some remained indeterminate, and why he thought others were acceptable methods for fluting Folsom points. These details generally concern formal attributes of the attempted replicas made, whether they were too fat, had too many undulations on the fluted surface, or have other blemishes. He was cautious in his language, leaving the door open on some techniques, but in his summary statements these cautions and techniques vanished, leaving only three surviving possibilities. And even this number was bothersome to Crabtree because his working assumption was that there was only one technique.

As described, Crabtree followed a procedure that began and ended with examination of archaeological materials, both for inspiration and final validation. Detailed analysis of the Lindenmeier epoxy cast established criteria for replication, or the replication target. From these observed attributes, he hypothesized various possible manufacturing techniques that could possibly culminate in this final result. These techniques were then tried or tested, and the attempted Folsom replicas were compared to the plastic specimen to see how closely they matched its diagnostic features. Some experimental points were found wanting and were rejected as replicas. This entailed rejection of the manufacturing procedures used to make them in the first place, and these were “eliminated” from the running as “The Technique” for fluting Folsom points. Conversely, pretty points that conformed to the replication target were acceptable, and their generative techniques were accepted as possibly “The One” used anciently.

Crabtree’s procedure for testing alternatives through test implications shares a vague resemblance to hypothetical-deductive thinking touted in the 1960s, but the resemblances are superficial. In Crabtree’s experiments, positive and negative results were accorded equal discriminatory power to inform about the viability or non-viability of techniques. If we take Crabtree’s truth claims and state them as formal arguments, from his point of view, some difficulties become apparent.

The negative case:
A. Folsom points are defined by attributes d, e, f, g, h, and sometimes m, n, and p.
B. Replicated points made with technique X have defining attributes a, b, c, l, and k, and they lack attributes d, e, f, g, h, m, n, and p.
BB. In other words, I was unsuccessful in making replicas with attributes d, e, f, g, h, m, n, and p with technique X, in my limited attempts to do so, under the conditions imagined, with the tools imagined, used in the ways imagined, and with my previous limited experience with the tools and technique.
C. Ergo, the attributes of replicated points do not match the attributes of Folsom points.
D. Therefore, technique X did not produce true replicas (under the specified conditions).
E. Therefore, technique X was not used to produce Folsom points.
The positive case:
A. Folsom points are defined by attributes d, e, f, g, h, and sometimes m, n, and p.

B. Points made with technique Z have attributes d, e, f, g, h, and frequently m, n, and p.

BB. In other words, I made replicas with defining attributes d, e, f, g, h, and sometimes m, n, and p with technique Z, in my limited attempts to do so, under the conditions imagined, with the tools imagined, used in the ways imagined, and with my previous limited experience with the tools and technique.

C. Ergo, the attributes of the replicated points closely match those of Folsom points.

D. Therefore, technique Z produced true replicas (under the specified conditions).

E. Therefore, technique Z may have been used to produce Folsom points.

If we fail to translate the experimental observation “B” into its constituent truth claims “BB” the arguments look stronger than they really are because they cloak the imaginings and leaps of faith involved. The negative and positive claims mirror each other and are grounded in “authority,” as mediated through empirical assessments of replication targets and experimental outcomes. It should be clear that Crabtree served as advocate, judge, and juror at his own trial, and ruled on his own acquittal. He was sole judge of what a Folsom point “is” and of any goodness of fit between experimental points and true Folsom points. He also decided the number of experiments necessary to confirm failure of experimental outcomes and their associated techniques. His was a lonely and commanding position, perhaps attributable to absence of colleagues at the beginning of a new endeavor who (he thought) could assess technical matters. Even ignoring these incredible weakness in auto-assessment of his claims, in the most charitable of all worlds his arguments would still not add up to a solid case.

The force of Crabtree’s argument derived from his imagined ability to eliminate possibilities, thereby narrowing the field of plausible techniques used to fluke Folsom points. But his assessments of negative results have consistently caused the most trouble, as subsequent knapping experiments demonstrate. Even IF granting that experiments with technique X failed to produce true replicas, one cannot justifiably infer that the same technique, performed by others with more or less experience and under similar or different conditions, would not produce replicas according to the criteria specified. In the first argument, “A” is the given parameter, “B” is an observation of experimental outcomes, with “C” necessarily following from “A” and “B.” In turn, “D” necessarily follows from “C,” but “E” does not and cannot follow from “D.” Jumping from specifics of experimental outcomes to a universal truth applicable to the Folsom past is not justified by any observations or entailed premises. The most interesting claims of the experiments, therefore, remain ungrounded.

Crabtree was completely justified in claiming that given experiments yielded specified results. This was simple observation and not subject to dispute. The problem came from generalizing on the basis of these experiments concerning a technique to its potential in all times and places. Getting from “did not produce” for an experiment to “was not used in the past” for the associated technique involved an illogical leap. Such a move would only be possible under very restricted conditions and after a comprehensive experimental program that tested all alternatives, and tested them fairly. Crabtree’s claim (# 2) to have “tried every conceivable method” stretches hyperbole. He derived inspiration for the range of possibilities tried from Holmes (1919), and his 17 combinations barely scratched the surface of the thousands imaginable.1 Making
positive claims from failed experiments is more problematic than making claims from confirming experiments. Crabtree could not logically eliminate any of the possibilities for fluting Folsom points he tested on the basis of his observations.

Crabtree did better with positive arguments of the second type, and his quandary of being unable to restrict possibilities to fewer than three techniques was an appropriate position that accords with the problems of equifinality and ambiguity (see Amick, Mauldin, and Binford 1989). He did not over-generalize his observations of “produce” to “was used”; rather he maintained the weaker, more appropriate claim of “may have been used” to produce Folsom points. Confirmation is firmer than un-confirmed ground. All replication experiments share this characteristic. After producing certain artifact forms with a technique, simple observation of the behavior and its consequences establishes grounds for certain knowledge; facile observation verifies that such forms can be produced under certain conditions with certain techniques. This is an analytical power absent in “negative” or non-confirming observations. As a classic case of inductive argument from particulars to generalities, no amount of failure with a technique suffices to establish the logical certainty of its failure in future attempts. We can only claim stronger or weaker suspicions that such will or will not be the case, depending on the range of observations subsumed under each inference.

At the center of these arguments lies the mental construct of a “replication target.” The care taken to specify this target, and of analyzing proposed replicas, determines the outcome of the comparative exercise and the strength of inferences claimed. What should be apparent in my simple attempt to make Crabtree’s arguments explicit is that truth claims for manufacturing techniques and their products ultimately rely on analogy supported by detailed comparative analysis of suites of attributes (see below). The notion that modern knappers devise experiments for deciphering aspects of the past remains problematic. What knappers really do is add to stores of modern knowledge that can be used in analogical arguments of archaeological phenomena. The distinction between the two classes of information and kinds of knowledge claims is critical. Knapping experiments do produce certain knowledge of procedures and ranges of material outcomes in the present, but they do not directly inform about the past. The relevance of current, secure knapping knowledge in interpreting archaeological remains depends entirely on conditions of justified analogy. And these can be no better than the rigor of comparative analyses and the appropriateness of the replication target.

Modern knapping experiments relate to ethnoarchaeology as means for increasing stocks of modern knowledge that causally link behaviors to material outcomes (see David and Kramer 2001). Information gathered under both approaches can be and is used to interpret similar-looking material outcomes surviving from various pasts. The two approaches are complementary. In contrast with ethnoarchaeological exercises, experimental archaeology cannot get at cultural meanings associated with behaviors because the cultural contexts of the present and past behaviors differ (see Dobres 2000). For its part experimental archaeology enjoys one huge advantage over ethnoarchaeology because of its unlimited potential to make novel observations of behavior and material outcomes. How to plan experiments to collect certain classes of observations, or to deploy such observations once obtained for interpreting artifacts, are separate epistemological problems (see Amick et al. 1989). Both experimental and ethnographic observations share equally the problem of relevance when it comes to their application to ancient remains. Eventual understandings of Folsom technology will depend on the range of experiments tried, their outcomes, and the
logical conditions for making strong and justified analogies to the past.

TOWARDS REPLICATING FOLSOM TECHNOLOGY

What could Crabtree have done to resolve the Folsom riddle? If his experiments in determining the technique for manufacturing Folsom points were indeed deficient, as I claim, and if his arguments were unacceptable on evidentiary and logical grounds, what can be done today to remedy the situation? My short review of Crabtree’s experiments suggests five areas in which modern replicators can make immediate and significant progress. First, we need more and better analyses and descriptions of archaeological collections that can serve in devising adequate replication targets and controls for comparative analysis. More care with experimental design and identification of critical variables will also be essential. Two critical issues concern the decision rules and protocols for evaluating the goodness of fit between experimental outcomes and replication targets and the structure and means of strong, justified arguments. The final issue involves questions of responsibility and public validation of experimental results.

Setting The Replication Target

Given the self-imposed limitations under which Crabtree worked, that he made as much progress as he did towards replicating Folsom points and discovering the ancient means of their manufacture remains a remarkable achievement. Most of the deficiencies of his experiments center on the inadequacy of his replication target and his knowledge of Folsom assemblages. The obvious solution to this weakness is to expand the range of observations involved to include whole assemblages, and not just additional finished points. Our interest ought to be in replicating the entire process for making Folsom points and not just the flutes and final edge retouch. Some care will be needed here to avoid possible debilitating assumptions of singularity. We should not assume a priori that all Folsom assemblages were precisely the same in all their details; assemblages from various times and places need to be studied with some care before deciding that they can be grouped for purposes of devising a better replication target. The analytical needs to monitor similarities and differences among assemblages must be balanced.

I use “target” here as a plural noun that includes many members, from finished Folsom points, fragments from broken pieces, and all debitage generated in their manufacture, use, and discard. By increasing the number of items involved in the target, it necessarily increases its discriminatory power as an eventual “experimental control” (see Flenniken 1984) during the comparative analysis phase of an experiment by increasing the number of parameters that have to be met. Improvements in delimiting the replication target must be both extensive and intensive. Extension moves us from solely considering final forms to including all initial and intermediate forms and their byproducts. We need to replicate whole populations of related items, and in their observed or inferred frequencies. Intension involves more detailed analyses of individual artifact types and classes. Crabtree mentions many possibilities here, as do all who followed him. Sollberger’s (1985) measurements of numerous points to establish the parameters of ancient points was particularly helpful, and many more analyses of this type must be done (see Chapter 9). It is not enough to have flutes, they need to fit certain parameters and exhibit certain kinds of features of fracture definition.

Much time at the Folsom Workshop Conferences was spent examining archaeological specimens and comparing their attributes to potential replicas. The general consensus was that no one present had yet made replicas that accounted for all
the forms of debitage and intermediate forms and/or the subtle features of finished points. Bob Miller’s observations were particularly interesting. Legally blind, he analyzed Folsom points by touch, and none of the replicas matched the smooth feel of the fracture path of the ancient points (it merits mention that weathered surfaces of archaeological specimens could have been a problem here). At both the population and individual artifact levels, we need to decide what the acceptable degrees of variation are. The target will be a statistical tendency, whether of the percentage of various kinds of flakes or manufacturing breaks, or the number of Folsom points in a sample that evince a particular characteristic. All of these concerns implicate issues of selecting samples and of grouping different samples and/or attributes. These decisions should be based on detailed and careful analyses.

A critical issue concerns sample size. Clearly, the larger the sample, the more confidence we can have in the abstracted statistics of the replication target. The sample should be greater than one (Crabtree’s sample size) and less than infinity. What is a practical solution? We need to take care to avoid conflation problems of working with assemblages that may have more than one technique and technology represented. Rules for sorting the archaeological collections and the grouping of like collections remain to be worked out. How does one sort through an assemblage to find the stuff relevant for an experimental program? Refitting work would be a great help in this endeavor. Considering assemblages from different cultural contexts will also be important.

There is more we can learn from Crabtree’s most avoidable error of having an inappropriate replication target. He was a consummate craftsman and meticulous worker, and he strived to do the best work possible. I think he set his standards too high, so high that they proved unattainable. Crabtree created idealized and ideal types in his mind to replicate, but none of them existed in any real world. He also chose final forms and lacked data on intermediate stages. He was fluting in the dark, and he missed by a mile—maybe.

**Designing Experiments And Determining Goodness Of Fit**

Experimental design and comparative analysis are separate analytical problems, but they are best considered together for practical purposes. The adequacy of experiments deserves to be monitored constantly by checking experimental outcomes against archaeological standards for the replication target (Amick et al. 1989). The utility of this procedure was clearly evident during both Folsom Workshop Conferences, with the continuous back-and-forth, or dialectic, consideration of experimental outcomes compared to artifacts in the displayed collections. Useful also was the running commentary of others engaged in parallel experiments. The dialectic movement between artifacts and experimentally generated pieces leads to more detailed observations and better perceptions of both, in trying to reconcile any similarities and differences. This comparative analysis occurs at two different levels. During experimentation, it is clearly inappropriate to sit down and do an exhaustive comparative analysis of outcomes and archaeological specimens. The dialectic I refer to happens on a more gross level of general assessments that provide some feedback and allow the experimenter to modify experiments in various ways to attain different results. More thorough analysis can be done later, and a more reasoned comparison made. The variables monitored, their means of measurement, and the comparison of the experimental and archaeological assemblages all need to be presented in sufficient detail that others can repeat the experiments and/or the analysis. Reporting remains a major issue (see below).

A question not raised frequently enough is how one knows that he or she has achieved an adequate replication. It is ridiculously simple to make such a
claim in print. In principle, validation of a replica should be straightforward. One should compare experimental outcomes carefully to replication targets and assess degree of fit. If experimental outcomes meet a certain standard, then they are replicas; if not, they are not. But what is the standard, how is it determined, and by whom? At the moment, the way things appear to work is that each experimenter bears total responsibility for the assessment. This is akin to having students write and grade their own exams and then award themselves degrees based on evaluations of their performance. This is a rather poor way to generate scientific knowledge. We need some system of peer review and verification—such as present in the 1920s in the original evaluation of the context of Folsom points (Chapter 22). Representatives of the scientific community had to see the Folsom point in context with ancient bison remains and then vouchsafe the finding for their colleagues. Two basic premises of the Folsom Workshop Conferences were that legitimate replication could not be performed without access (1) to adequate archaeological collections or (2) to the opinions of other experts. Evaluation of experimental results, as with other scientific claims, is necessarily a social affair involving a community of scholars. It is not enough that a replicator broadcast claims for having made legitimate replicas; such a claim must necessarily be verified by independent evaluators. Crabtree's solipsistic approach was clearly inadequate. Unverified reports of replication by individuals cannot count as serious claims.

In my view, experimentation is a commitment to a process of discovery and verification rather than the specifics of any given experiment. The previous ragged record of "experimentation" with Folsom points more than adequately demonstrates the need to involve a community of scholars in the process of evaluation. Critical are the identification of interesting issues and variables. As demonstrated by papers in this volume, isolation of critical variables is a key move in designing experiments (see Amick et al. 1989). Of particular importance are clear descriptions and analyses of the outcomes of various experiments as evaluated against archaeological targets or standards. The "success" of an experiment is assessed through the comparative study of its results. But the rules for evaluating experimental knapping outcomes remain to be worked out among the interested coterie of researchers. What constitutes a good or poor fit in experimental outcomes? How much difference do we tolerate before deciding that the population of objects created experimentally differs significantly from the populations of similar objects, known archaeologically? Of equal importance are issues of experimental rigor, measurement, and repeatability. How many times was a given experiment successful or not? As argued here, one serious weakness of Crabtree's approach was that he sealed off avenues of research prematurely, with insufficient experimental effort. His actions were understandable from numerous points of view, but they do not constitute good procedure for a replication programme (see Clark 2002).

Consider practical issues of experimental knapping versus "lab knapping." In the lab, with mechanical devices, it is possible to monitor carefully, and control for, a range of variables. In actually trying to test a procedure with flesh-and-blood knappers, often several dozen variables will be altered at once for different experiments. Thus, it is not possible to isolate easily the cause or causes of unsuccessful experiments, however defined. For example, with a simple experiment of trying to flute a Folsom point using an antler baton (see Chapters 12 and 15), if an experiment is "unsuccessful," do we attribute it to the size of the baton, its overall shape, the shape of the portion contacting the preform, the force of the blow, the angle of the blow, the density of the antler or the billet, the holding position of the preform, the shape of the preform, the care taken
with the nipple for fluting, knapper indigestion, distraction from spectators, dust in the air, or other factors?

Lab experiments are clearly very useful for getting a handle on the range of possibilities. The new technique of “computer knapping” with Finite Element Analysis described by Tony Baker (Chapter 11) is also a very fruitful and helpful avenue, as his ongoing collaboration with Bob Patten has shown (Chapter 15, Patten 2003).

It is unrealistic, however, to suppose that all experiments can be attempted a sufficient number of times, or that all variables can be tightly controlled in all experiments, to establish the full universe of possibilities. The various kinds of knapping experiments (live, lab, and computer) have to confront practical issues of costs and benefits. Crabtree relied on his experience to eliminate some possibilities as quickly as possible in trying to find the technique for fluting Folsom points. All experimental programs will have to devise a list of priorities, and an hierarchy of variables to be tested, so the best results can be achieved in a pragmatic program that evaluates the most probable possibilities with affordable efficiency. Discussions on all of these fundamental issues is still required. What are the key variables, and what is their probable order of importance in their influence on experimental outcomes, and by implication, on the output of ancient knappers? How can we adequately bracket some critical variables? What kinds of simplifying substitutions can we make, for example, in the types of materials worked and in the materials used to work them? Is it legitimate to use copper? Glass? Vises? Why, or why not? The control of variables is always more or less—being close may suffice to answer some critical questions. How close do we have to be? And how much lumping of specific variables into more general classes can we do? We need to match questions with experiments and with some ability to make generalizations from specific experiments. For

some issues, such as the time it took to make an artifact, for example, just getting in the ballpark provides the proper degree of magnitude to make a strong inference to aspects of ancient behavior.

At the moment, there has been insufficient replication attempts to get a good read on possible problems involved in working out some of these issues. Ken Rozen (Chapter 14) reports the most thorough series of experiments of which I am aware, and as he readily admits, his experiments barely scratch the surface. Progress in the field will build from many efforts such as his, and others reported here (Chapters 11–15). In deciding these issues, we can look to the hard sciences for experimental design, protocols, reductionism and isolation of variables, reporting procedures, and procedures for validation—both through repetitions of experimental outcomes and through peer review of results. Truth claims cannot exceed the initial rigor of experiments, care in observations, measurement standards, comparisons, or adequacy of peer review.

Arguing Responsibly

Scientific knapping is hands-on philosophy, and those engaged in it cannot escape the responsibility of deep and serious thinking. Logic should be a familiar hammer in one’s toolkit. In terms of truth claims made for various replication experiments, few have bothered sufficiently with requirements for logical argument. So the need for, and the potential of, reevaluating previous claims in this light remain promising. Logic enters in at all stages of the experimental process, from isolating variables to evaluating the goodness of fit among various classes of objects.

I have already reviewed logical problems with most replication experiments, as exemplified by Crabtree’s Folsom experiments. Most knapping experimentalists have worked by the process of elimination and have overstated their negative results. It is much more difficult to reject knapping
possibilities based on experimental outcomes than it is to accept experimental outcomes as possibilities for the past. As with all tests of hypotheses, there are two possible types of errors here. We can reject a hypothesis when it should be accepted, or accept it when it ought to be rejected. To be able to do either in the proper way requires that we work out standards for evaluating goodness of fit, and parameters for acceptance or rejection of experimental results based on the score assigned for the goodness of fit within ranges of probability. As with the knapping experiments, these logical procedures need to be consistent and repeatable.

What ought to be the criteria for goodness of fit? What attributes ought to be considered, at what scales, with what variance, and with what percentage of correspondence? How many duplications of the experimental results are required? Are there any independent controls on reconstructions of manufacturing sequences and technology? What is the sociology of verification? How many witnesses are needed?

Replication is tough work, and as a profession we have been doing it poorly and unconsciously—philosophically speaking. Replicators consistently confuse personal belief for public knowledge, and perhaps experience for experiment. If we phrased our results as beliefs rather than knowledge claims, we could get out from under the philosophical burden, but to do so would reduce the enterprise to irrelevancy. If knowledge really can be derived from modern experiments, its production and validation ought to be our goal. Validation procedures provide a way for sorting out personal impressions from legitimate knowledge claims.

Validating Results

Validation of knapping results involves descriptions of experimental outcomes and artifacts, detailed comparisons between the two at various levels or scales of observational detail, logical argument and inferences, and public access to all of the above. The outcomes of each step of the analysis require their own validation, and this requires the approval of the interested community of scholars. It is not practically feasible to involve many scholars in the process before publication, thus publication becomes the principal means for establishing a record that can be independently evaluated. But most experiments are not presented in sufficient detail that would allow for independent assessment by those not present at the experiments. Rozen's (Chapter 14) efforts set the standard of adequate reporting that we all should strive for. Putting all details on record is obviously a tedious process, and not all publication outlets will be appropriate. Alternative ways of putting the experiments and their outcomes on public record will be required. Some of the avenues currently being explored for electronic journals will probably well serve this purpose.

Costs of publication and writing difficulties are the least of our concerns, however. Of critical importance are the standards of descriptive adequacy. These obviously will change through time for any given technology once more becomes known about it. For example, many of the attributes described by Patten (2003, Chapter 15) for Folsom points represent new views of the critical features of these points, and these basic observations have not been taken on most collections. One of the benefits of interactive replication experiments as a way of "seeing" is that it helps sort out the traditional variables that we rely on for basic description from those that are actually important for addressing specific questions. One such feature is the "gap" distance between the plane surface of the preform and the "set-back" nipple for the Folsom preform described by Patten. It looks like an unremarkable thing, but its implications are huge, once its function in thwarting unsuccessful fluting is understood.
Descriptive standards ought to include standards for adequate illustration as well. Another issue concerns the curation of experimental collections. Once one has trudged through the labor of describing in detail serious experiments, then the collections described ought to become part of the public record so others can actually examine them. In one sense, they have greater value than most archaeological collections because there is a written, and perhaps visual, record of how they were made. Some of the knapping tools used also ought to become part of this curated collection. One reason we chose to hold the Folsom Workshop Conferences in Austin is that Mike Collins and I anticipated that more experiments would be done and that the resulting collections could be housed by the Texas Archaeological Research Laboratories there. This did not happen. But it would be nice if there were a public repository for the results of experiments so that other analysts could rely on them as resources, and it would be well if these were centralized in one place rather than being scattered among different institutions and private collections. Curating experimental collections with adequate documentation is the ultimate way to validate experimental results. To his credit, Crabtree put all of his serious experiments on file, and they can be studied by others. We need to be more serious in following his example.

EXPERIMENTS, ANALOGY, AND STRONG ARGUMENTS

Preceding matters have concerned Crabtree's preliminary experiments, embedded logic, and view of the Folsom problematic. With analogy as the fundamental logical link connecting modern experiments and ancient Folsom points, it follows that replication experiments cannot establish with complete certitude the technique or techniques used to manufacture these points. But it is possible to establish a strong presumption that approaches near certainty, and this should suffice for all interpretive purposes. Such levels of confidence will require proper use of analogical arguments. These can be advanced from both the archaeological side and the experimental side.

Legitimate analogical argument requires assessments of correspondence between ancient remains and modern remains of known derivation. For Folsom matters, the primary way to advance understanding of its manufacture will come from more and better descriptions of Folsom points and their related byproducts. This was a principal goal of the Folsom Workshop Conferences and one of its enduring accomplishments. The screaming weakness of Crabtree’s experiments was his ignorance of the artifacts themselves, and this crippled his imagination about the manufacturing sequence and plausible techniques. Detailed knowledge of artifacts and their obvious and subtle features at an assemblage level is absolutely essential for determining how Folsom points were made. Careful analyses and descriptions of adequate archaeological collections, in turn, serve both as replication targets and as criteria for assessing experimental outcomes.

Matching modern and ancient material outcomes will depend on justified, logical arguments, and these will involve practical matters such as decision rules for distinguishing legitimate experimental replicas of ancient artifacts from fabricated pretenders. Production of legitimate replicas is the basic entrance requirement for evaluating a technique. Returning to Crabtree’s procedure for replication, the first step of analysis of archaeological materials necessarily involves several logical operations. Descriptive standards are critical but insufficient for experimental purposes. Analyses of collections inspire ideas about technological sequences and possible behavioral connections among the various fragments. Hypotheses about manufacturing sequences and possible techniques are necessarily based on induction. As with analogy, induction can-
not lead to absolute certainties (only stronger or weaker conclusions) because inferences are ampliative; they exceed information stated in the premises (Honer et al. 1996). Stated as hypotheses, however, inductive inferences can be tested through modern experiments. The careful comparison of ancient and modern artifacts, then, is more precisely considered as an evaluation of inductive inferences from archaeological artifacts, allowing analysts to assess the likelihood that various techniques were used in their production. Some ideas concerning manufacture can be evaluated independently through refitting and recovery of fragments of manufacturing tools or residues.

CONCLUDING REMARKS

Because of practical constraints, my treatment here of Crabtree’s Folsom experiments has only skimmed the surface of possibilities. My original intent was to find, classify, and analyze each truth claim in his Folsom paper, and related articles, and evaluate the physical evidence and assumptions behind each one. I wanted to recover the patterns of his arguments and to evaluate their logical merit. I also intended as part of the same project a lengthy treatment of implicated philosophical matters such as realism, empiricism, perception, objectivity, and analogy, to name the obvious ones. This master objective remains unrealized; it clearly will require years’ more effort. The magnitude of the task exceeds present space requirements. For example, as one assignment for an ancient technology class, while I was still helping plan the first Folsom Workshop Conference, I had each student work through Crabtree’s article and list his sequential truth claims and the assumptions that each claim implied. Ignoring scores of trivial declarative sentences, there are still over 100 significant truth claims and critical assumptions—too many to include in a chapter-length treatment here. On the other hand, the philosophical issues related to knapping experimentation, observation of outcomes, comparative analysis of collections, and possibilities of legitimate inference by analogy about the past are a veritable quagmire that I am ill prepared to traverse. I avoided these unchartered swamps here by curtailing my original expectations of making a thorough analysis. I have mined, instead, only a few obvious lessons from Crabtree’s Folsom failures.

Evaluation of Crabtree’s general claims exposed three problems areas: (1) delimitation of the replication target, (2) evaluation of experimental outcomes, and (3) validation of experimental results. These need not be limitations for any future experimental program because all three issues can be readily solved by increased access to archaeological collections and peer review. My specific interest in promoting a Folsom workshop, beyond learning more about this fascinating technology, was to have the workshop itself be an experiment in the sociology of knowledge, by which I mean the social processes within the discipline that promote valid truth claims about the past. It is still too early to report on this meta-experiment, although the numerous successes of individual efforts reported in this volume, and elsewhere, bode well for the workshop-conference concept as a means for producing valid knowledge claims.

Crabtree’s basic logic for progress in replication experiments was through elimination and validation. He appears to have come up with the process of elimination on his own. The earliest knapping experiments (see Johnson 1978), at least up to H. Holmes Ellis’s (1940) work, focused on the “feasibility” of different techniques for duplicating ancient objects. Crabtree raised the stakes and thought it possible to know the specific technique used to manufacture specific objects. I share his optimism but think it a much more difficult inferential task than supposed.

We can perhaps trace the origins of the strong epistemological claim of “impossibilism” in Crab-
work to the “eolith” controversies at the turn of
the 20th century (see Grayson 1986 for a good sum-
mary). The emphasis in these experiments was to
define the limits of what Mother Nature could do
by way of flaking flints. Coupled with notions of
design and repetitive motion, the plausible possibilities
of nature are insufficient to account for the
repetitive, precise, applications of force evident in
the scar patterns on some tools—thus they are pre-
sumed to be un-natural or human flaking patterns.

Clearly, Folsom points were made by humans, but
there was a time not far distant where this was not
obvious to many people. Precisely how humans
made these points, beyond a few generalities of
reduction sequences that can be read from flakes
or chunks (another benefit from early experiments),
remains an open question, and perhaps still an inter-
esting and potentially revelatory question as well
(Chapter 20).

At some level, the eolith experiments, coupled
with observations of natural processes, established
the soundness of “impossibility” as a means of in-
ference. No one today would question that Folsom
points were made by humans. The very question
appears absurd. This fact suggests that at some level
of analysis, certain knowledge concerning manufac-
turing techniques may be had if the experiments are
extensive and thorough enough. What these condi-
tions will prove to be has not been determined. In
pondering the question, it would be useful to reex-
amine the earlier debates on eoliths to see whether
our current confidence in the conclusion of the
controversy is really well supported or merely
another disciplinary deceit of changing scientific
dads. In any event, the requirements for making
positive claims of what cannot be done based on
knapping outcomes are much more stringent than
the alternative of making positive claims about
what can be done.

Determining the limits of natural knapping was
a far simpler task than proscribing the limits of any
human technologies. The oldest replication work,
going back to John Evans’s (1872) master publica-
tion, done in conjunction with observations of Bran-
don flintknappers, followed the dictates of “possibil-
ism.” Early scholars made claims about how ancient
artifacts could have been made, but they did not
exclude possibilities by making claims on how they
could not be made. One cannot prove a negative, but
it is feasible to delimit the possible, thereby roughly
bracketing probable ranges of the impossible. Con-
trary to Crabtree’s occasional rhetorical excesses, we
can never claim to have exhausted all possible knap-
ning techniques and combinations or to have
accorded each individual combination an adequate
chance of having been executed skillfully. Therefore,
we cannot claim as much for our knapping failures as
Crabtree did for his. A major failing of Crabtree’s
Folsom experiments is that he misunderstood the
significance of his failures. Future experimentalists
can easily avoid this error, and others.

To summarize, my brief review of Crabtree’s
Folsom experiments indicates a few obvious ways to
improve replication experiments that we can imple-
ment, even without fully mastering the remaining
metaphysical and logical issues. (1) We can better
analyze and describe archaeological collections in
all their subtle details; and these will serve as repli-
cation targets as well as experimental controls. (2)
We can devise more extensive replication experi-
ments that methodically control for different vari-
ables. (3) We can adopt methods for comparing the
results of experiments to archaeological targets. (4)
We can institute a system of peer review to pass
judgements on proposed replicas. (5) We can work
to establish more rigorous criteria for accepting
pieces as replicas; we lose nothing by setting a high
standard as long as it is based on observations of
aboriginal artifacts. These have to be specific to
each artifact class and technology. (6) We can
determine archaeological signatures of technology
at the assemblage level. (7) Through a program of
experimentation, we should be able to establish what is possible and what appears implausible; we probably can not establish impossibility. (8) In cases of equifinality, we need to search for independent evidence among other artifact classes for the ancient techniques employed (i.e., knapping tools and holding devices). (9) We need to evaluate hidden assumptions concerning replication, even the obvious ones such as privileging parsimony. (10) And, we need to state clearly the hypotheses that we consider the most plausible and to lay out their test implications in such a way as to promote further testing. Experiments will advance faster if we purposely devise them so failure is a clear option, and if failures have predetermined implications for delimiting the range of possible or probable techniques used in the past.

NOTES

1. Crabtree did 17 different experiments, and they correspond precisely with those described by Holmes (1919) in his famous treatment of aboriginal stoneworking, but they fall short of the 22 different techniques attempted by H. Holmes Ellis (1940). All fall well short of the universe of the possible. The number of possible significant experimental combinations can be quickly appreciated by a simple exercise with ballooning numbers. Suppose we are dealing with 4 possible techniques, 4 possible tools that can be used with any of the techniques, 4 ways to hold each tool, 4 ways to immobilize preforms for fluting, and 4 ways to prepare the preform and the fluting surface and nipple platform. Even forgetting all other necessary variables, and additional variants of those specified, if each variable were independent of the others, this would give us $4^4$ possibilities, or 1024 combinations. Throw in a few more variables and variants, and this quickly moves to tens of thousands. Of course, not all combinations are possible or feasible, so the total number of possible combinations would be a从业人员 of infinity. Clearly, not all variants would be equally plausible, but sorting through plausible combinations is a separate issue. How much time ought to be accorded experimentation with each variant is also a separate but important issue. If we had good reasons to narrow the field of the plausible to 5000 combinations and only conducted 3 experiments for each, we would still be strapped with several lifetimes of work. The actuality of things is a bit more complicated because standard theorizing does not deal with critical micro-variables of whether or not one's antler punch is too worn, whether or not one's hammerstone is developing a facet and, therefore, the effective working angle involved in knapping the stone is changing during the process of manufacture, and ad infinitum. Serious replication is not for the faint-hearted. The simple escalating mathematics involved suggest further that it should not be the responsibility of individuals. So much work is implicated that replications ought to be conducted by coordinated effort of many experimenters, or what I call experimental “programmes” (Clark 2002).

ACKNOWLEDGMENTS

I thank Dan Amick, Joel Janetski, Bob Patten, and Gene Titmus for reading and commenting on an earlier draft of this paper and helping me improve it. The issues discussed here benefitted from the generosity of all those who attended the Folsom Workshop Conferences and who allowed me to watch their work, examine their archaeological collections, or engage them in conversation about replication experiments and Folsom archaeology. I greatly esteem all their kindnesses and friendships.