ARTICLE

HOW CAN ARCHAEOLOGISTS MAKE BETTER ARGUMENTS?

Michael E. Smith

Michael E. Smith (mesmith9@asu.edu) is Professor in the School of Human Evolution and Social Change at Arizona State University.

> The fundamental question of all serious fields of scholarly inquiry [is]: How would you know if you are wrong? [Haber 1999:312]

Thanks to advances in methods and fieldwork, archaeology today seems poised to make major contributions to knowledge that extend beyond traditional archaeological and anthropological concerns. The recent identification of a series of "grand challenges for archaeology" (Kintigh et al. 2014) illustrates the applicability of our data and findings to major research questions across the human and natural sciences. But at the same time, standards of argumentation have declined, particularly in the anthropological archaeology of complex societies. As a result, increasing numbers of our published interpretations and explanations are weak and unreliable, and our empirical data fail to contribute to the accumulation of a solid body of archaeological knowledge.

Over the past decade, I have explored other social science disciplines by reading and interacting with colleagues on interdisciplinary research projects. I have been embarrassed at the frequent lack of rigor in archaeological argumentation in comparison with other social sciences. But at the same time I have been pleasantly surprised at the relevance of methodological and epistemological works in these non-anthropological fields. Both graduate training and the published literature in archaeology tend to neglect such issues. I did a brief nonsystematic survey of the content of graduate classes on theory and method at U.S. universities, and I found very little concern with the topics of argumentation and explanation beyond the occasional treatment of analogy. Yet these are fundamental topics in the many textbooks on social science methods available today (e.g., Abbott 2004; Booth et al. 2008; Gerring 2007, 2012).

In this paper, I identify a number of problems with current practices of archaeological argumentation and explanation, and I suggest some solutions that have proven effective in the social sciences. I start from the position that archaeology is a science in the sense that knowledge is responsive to evidence; claims are exposed to challenge; findings should be internally coherent; and arguments should be judged on the basis of explanatory power, generality, simplicity, and replicability (Gerring 2012; Wylie 2000). Given the constraints of this paper, I can cite only a small number of relevant sources; I focus on the most important references, and these can be checked for further citations.

The Structure of Arguments

Lewis Binford and other archaeologists affiliated with the New Archaeology movement gave a fair amount of attention to the structure of archaeological arguments and the nature of explanation. Unfortunately, many of their formal arguments were unsuccessful because they accepted a faulty view of explanation, the "covering law model" of the logical positivists. This kind of explanation had been shown to be inappropriate for the social sciences even before it was adopted by archaeologists. When archaeologists were unable to construct useful covering law explanations, they began to neglect epistemological issues, a trend that has affected work in both the scientific and post-processual approaches. Yet in the other social sciences, such questions have long been staples of methodological concern and student training. I draw from that literature in discussing the structure of archaeological arguments.

How Would You Know If You Are Wrong?

The epigraph from economic historian Stephen Haber gets to the heart of effective argumentation. This precept is discussed in just about every textbook in social science methods (e.g., Gerring 2007:74-75; Ragin and Amoroso 2011:39). It is a relaxed version of Karl Popper's well-known concept of falsifiability as a criterion for a scientific argument. Popper had a strict concept of hypotheses that could be falsified with a crucial experiment. Subsequent philosophers of science showed that his scheme was too rigid for the social sciences, but the basic idea that one must be able to tell when one is wrong remains a critical foundation for empirical research. Here is how sociologist Andrew Abbott frames the issue:

> It is surprising how many researchers—even graduate students in their dissertations—propose arguments that can't be wrong. For example, research proposals of the form, "I am going to take a neo-institutionalist view of mental-hospital foundings" or "This paper analyzes sexual assaults by combining a Goffmanian account of interaction and a semiotic approach to language" are not interesting because they do not propose an idea that can be wrong. They boil down to classifying a phenomenon or, seen the other way around, simply illustrating a theory (Abbott 2004:216).

Post-Hoc Arguments

The post-hoc argument is a problematic form of explanation. Lewis Binford (e.g., 1981) discussed problems with this procedure, which he called "post-hoc accommodative argument." He was referring to an interpretation that is applied to the data and findings once the research activities are complete. The problem with post-hoc arguments is that they can't be shown to be wrong. The analysis is done, and the post-hoc interpretation cannot be disproven without another round of research. We can all dream up numerous alternatives to explain (or explain away) any set of findings. But without some form of testing, post-hoc arguments serve to introduce potentially faulty or misleading interpretations into the literature.

In experimental scientific fields, post-hoc arguments are considered unacceptable and even unethical. Psychologist Norbert Kerr (1998) labels the practice "HARKing" (Hypothesizing after the Results are Known), and he presents a list of 12 methodological and ethical problems with the practice. These include non-falsifiability, the encouragement of data fudging, and the promotion of narrow, context-bound theory. In experimental psychology, post-hoc arguments are seen as the scientific equivalent of the farmer who paints bulls-eyes around the bullet holes in his barn in order to show his superior shooting skill. In historical disciplines in which narrative explanations are common (such as archaeology), post-hoc arguments are less pernicious because research is not structured in terms of discrete individual experiments. Explanations are more commonly phrased as inductive, rather than deductive, conclusions. But John Gerring (2007, 2012) and others have proposed that most social science research (including case history and historical disciplines such as archaeology) can in fact be viewed as a form of the experimental method, and thus post-hoc arguments are best avoided if possible. Below, I discuss alternatives to this form of argument.

An Idealized Argument Structure

The faulty procedures outlined above involve the failure to test one's arguments. Archaeologists sometimes claim that hypothesis testing is inappropriate because our field is not an experimental science. There are two responses to this claim. First, many non-experimental sciences, from geology to astronomy, incorporate testing into their research designs and arguments. Second, the concept of experiment can be broadened to include testing procedures in non-experimental sciences, as mentioned above.

1. **What is the claim?**
2. **What reasons support the claim? (base claims on reasons)**
3. **What evidence supports the reasons? (base reasons on evidence)**
4. **Acknowledge alternatives/complications/objections**
5. **What warrant or principle justifies connecting the reasons to the claim?**

Figure 1. Idealized argument structure (from Booth et al. 2008:Chapter 7).

Booth et al. (2008) present a model for the construction of rigorous arguments that is designed to be applicable in all fields of empirical scholarly inquiry. Because of its generality, the model needs some modification for fields like archaeology with special data requirements. The model is outlined in Figure 1. An argument begins with a *claim*. A claim is equivalent to a hypothesis at the point when it is being supported or refuted, after gathering the relevant data (and not a hypothesis at the time of its initial formulation in advance of research). Claims or hypotheses are based on *reasons*. Reasons and claims are connected by principles of logic that should be acceptable to the community of researchers. The reasons are supported by empirical *evidence*. These three steps comprise the basis of argumentation in the framework of Booth et al. But for an argument to be convincing, two additional steps are required.

First, one should *consider alternative claims*, possible objections, and complicating factors. Without such consideration, arguments can sound limited and forced. Few if any arguments in history or the social sciences are completely water-tight, and acknowledging alternatives and contrary facts serves to contextualize a claim and give it greater strength and believability. Second, arguments need a *warrant* or principle that justifies the links among claims, reasons, and evidence. The topic of warrants is the weakest part of Booth et al.'s model, and it is an area where archaeological arguments have distinctive properties.

The two primary types of archaeological warrant are theory and comparative data. Arguments should be justified on the basis of one or more theoretical principles, and they should not violate accepted theoretical precepts. Arguments should also be justified by citing comparative data that establish at least the plausibility of the reason and claim. Comparative data are normally invoked in two areas of archaeological epistemology: discussions of analogy (usually termed "ethnographic analogy") and explicit comparative research. But in fact comparative data can and should be invoked in many archaeological arguments, whether or not one is making a formal argument by analogy. Such comparative data does not have to be ethnographic; it can be the results of other archaeological research, data from historical research, or findings from modeling or other methods.

Booth et al. (2008:169) discuss two types of arguments. They note that researchers tend to give more credibility to claims that are backed up with reasons based on evidence than to claims that are inferred from a reason and a warrant. The latter is an argument based on general principles, not on evidence. This point brings up an important consideration that is too often lacking in archaeological arguments: specification of the strength of the evidence and the strength of the argument. Some arguments are much stronger than others, and it is important to distinguish arguments that are weak and speculative from those that are strong and well supported.

Pitfalls to Avoid in Citing Comparative Data

I have been dismayed at the growing use of a faulty form of argument, particularly in the archaeology of complex societies. One begins with a discussion of high-level, abstract social theory (e.g., practice theory, materiality, post-structuralism, actor-network theory). Then one presents one's archaeological methods and results, with little reference to social theory. Finally, in a concluding section, one claims that the theory explains the data, or perhaps that the data illustrate the theory. For reasons discussed below (see also Smith 2011), this is a particularly weak form of argument. There is no testing, and linkages between theory and data are rarely operationalized in an explicit fashion. Most archaeologists recognize this as a weak argument, and some will therefore employ two kinds of faulty warrants to support their claims: ad hoc analogies and empty citations. While these devices may appear at first glance to give greater support to the argument, in fact they do nothing of the sort.

Ad Hoc Analogies

By the time Alison Wylie published the definitive paper on analogy in archaeology (Wylie 1985), this method had become embedded in archaeological practice. Although work on hunter-gatherer bone use, termed "actualistic" research, continued as a field of active methodological debate (Binford 1981), explicit attention to analogy receded from much of the archaeological literature. Perhaps graduate programs stopped teaching the method of argument by analogy; it now seems that poor and weak analogical arguments are almost as common as rigorous and strong arguments.

As argued explicitly by Binford (1967), and then formalized by Wylie (1985), an argument by analogy is a form of inductive logic. Wylie's criteria for evaluating the strength of an analogical argument very closely match the criteria for evaluating inductive arguments as described in textbooks on logic (Copi 1982:397–400). The methodological guidelines are quite simple. One creates an analogy or comparison with known human activities or patterns in order to interpret one's archaeological finds. The analogy is then treated as a hypothesis for testing, not in the sense of a definitive or crucial test, as in Popperian falsification, but rather in the form of the evaluation of alternatives. As discussed by Wylie, such evaluation proceeds in two directions: source-side research (increasing the number of cases and/or the quality and precision of the ethnographic, historical, or other analogues), and subject-side research (better specification of the archaeological case or cases).

Instead of following these simple and well-known guidelines, many authors today invoke analogy by citing one, or perhaps two, analogical cases from anywhere in the world that seem somehow related to the argument at hand. I refer to these arguments as "ad hoc analogies." There is little consideration for sampling or formal comparison. Ad hoc analogies provide no support at all for the argument at hand. The fact that some human group somewhere in the world did something vaguely similar to what you are claiming for your archaeological case does not in fact support your claim.

Empty Citations

> Empty references are references that do not contain any original evidence for the phenomenon under investigation, but strictly refer to other studies to substantiate their claim. Other authors subsequently use these empty references to substantiate their claims rather than going back to cite the original source (Harzing 2002:130).

Information specialist Anne-Wil Harzing was puzzled when the accepted wisdom in a field she knew was at odds with the data as she understood them. She created a dense chronological diagram of published works, showing which sources supplied data and who cited whom. When she removed the empty citations (works cited by other scholars but not contributing any new data), the diagram was considerably thinned out, and the studies that remained in fact supported an interpretation that was

the opposite of the accepted wisdom. She discovered that authors who wanted to lend an aura of support for a weak interpretation tended to cite many sources, and these included many empty citations. Similar cases abound in the social sciences.

Empty citations are unfortunately quite common in archaeology today. Authors boost the apparent level of support for their claim by citing a number of works that discuss similar claims, perhaps in different regions or time periods. If the general proposition is plausible and widespread, then one's particular claim may have greater validity. But if those references are empty citations and not citations for data, any increased support for one's claim is illusory.

Productive Frameworks

In this section, I describe two explanatory frameworks that have been employed successfully in many scientific disciplines, including archaeology: strong inference and natural experiments.

Strong Inference (Multiple Working Hypotheses)

Strong inference is the term used for a process of hypothesis testing that originated with Francis Bacon. The standard description is an influential paper in the journal *Science* by physicist John Platt (1964). His model has four steps (Figure 2). This approach is often called the method of multiple working hypotheses, citing a paper from 1890 that was reprinted in *Science* after Platt's article became popular (Chamberlin 1965). Although Platt made a number of claims that were later disproven (e.g., that scientific disciplines that employ strong inference advance more quickly than others), there is no doubt that his paper had a significant effect on scientific methods in many disciplines. In the words of biologist Rowland Davis (2006:244), "Platt imparted to many natural and social scientists an ambition to test hypotheses rather than to prove them."

The method of strong inference was championed by Chamberlin and Platt as a way of avoiding a researcher's tendency to become too strongly attached to a favored model or hypothesis, which can lead to efforts to merely confirm rather than disconfirm hypotheses. If the data merely confirm one's pet theory, then one has not significantly advanced knowledge. We can all think of archaeological examples here. For the social sciences (and archaeology), the only adjustment needed is a relaxing of the requirement for "crucial experiments." The two big advantages of strong inference are its emphasis on testing and the ability to know when you are wrong. One can evaluate multiple hypotheses, and, even in the absence of crucial tests, the last hypothesis left standing is most likely to be the correct one. The method of "inference to the best explanation" is a closely-related approach that is also very useful in archaeology (Fogelin 2007).

1. **Develop alternative hypotheses for the case at hand.**
2. **Devise crucial experiments to eliminate all but one hypothesis.**
3. **Perform the experiments.**
4. **Start the process over.**

Figure 2. Strong inference (Platt 1964).

Natural Experiments

A natural experiment, or quasi-experiment, is "an observational study that nonetheless has the properties of an experimental research design" (Gerring 2007:216). That is, cases are divided between a "treatment group" and a "control group," and while their group membership is not assigned randomly, it is effectively random with respect to the outcome of interest. Jared Diamond and James Robinson (2010) assemble a number of natural experiments in society and history, including an archaeological case (by Patrick Kirch). A big advantage of experimental methods, including natural experiments, is that they permit stronger inferences about causality.

A common natural experiment framework examines the effects of strong outside forces (perturbations) on a group of similar or related social contexts. I have employed this strategy in examining the effects of imperial conquest on households in several excavated communities in central Mexico. In one case, several very different sites were subject to the same perturbation (conquest by the Aztec empire), resulting in divergent consequences; this suggests that local conditions were of greater or equal causal import than Aztec conquest in generating the observed changes. In another case, two settings were each subject to successive conquests by the Aztec and Spanish empires, resulting in parallel outcomes: few changes after the first conquest, but major destruction and transformation after the second. These results suggest that the observed changes were caused more by the nature of the empires than by local conditions at the sites (Smith 2016).

Theory and Explanation

High-Level Theory and Middle-Range Theory

As I discuss in more detail elsewhere (Smith 2011), there is an epistemological hierarchy linking different levels of theory, data, and the empirical world (Figure 3). Although some post-processual archaeologists have denied the existence of distinct levels of

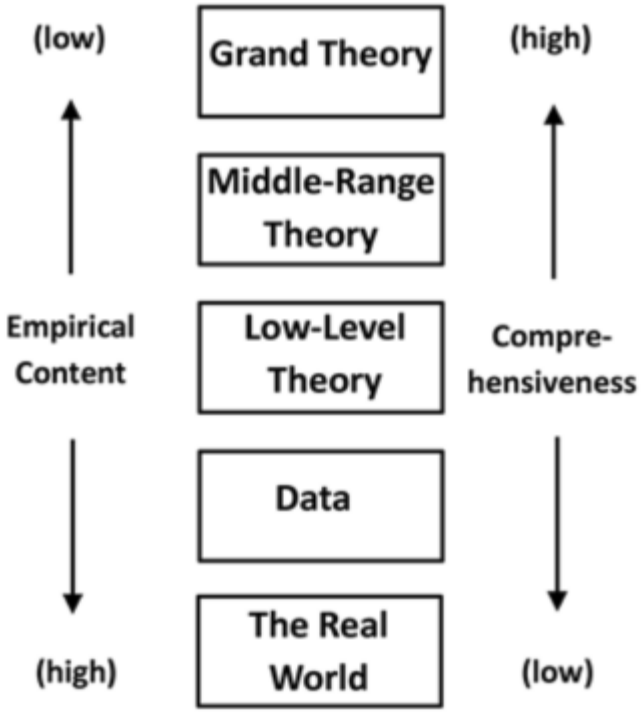

Figure 3. Epistemological hierarchy (based on Smith 2011).

theory (see Smith 2011:170), this is in fact a standard model in the social sciences (Abend 2008). High-level, or grand theory, is abstract and philosophical. Practice theory, post-structural theory, or materiality theory, for example, are broad and comprehensive visions of the world, but difficult or impossible to apply directly to explain individual cases without additional bridging concepts. Those concepts are supplied by middle-range theory. This refers to a widely accepted approach first articulated by sociologist Robert Merton in the 1940s. Later, Lewis Binford adopted this same term for a very different archaeological concept of formation processes. Although Raab and Goodyear (1984) tried to disentangle Binford's usage from Merton's concept, archaeologists have been slow to take advantage of the latter. According to Merton, middle-range theories:

> lie between the minor but necessary working hypotheses that evolve in abundance during day-to-day research and the all-inclusive systematic efforts to develop a unified theory that will explain all the observed uniformities of social behavior, social organization, and social change (Merton 1968:39).

In the social sciences today, considerations of explanation and causality are closely bound up with middle-range theory. High-level theory, on the other hand, is preferred by social scientists of a more philosophical bent who want to interpret society in abstract terms.

This epistemological hierarchy is highly relevant to the success of archaeological arguments. High-level theory cannot be tested directly, and it offers few clues to explain specific empirical facts. If one accepts the ideas of strong inference and the necessity of testing archaeological propositions, then high-level theory is of little use in devising arguments that work. One needs middle-range theories that can be operationalized and tested, and rejected when appropriate. Scholars who use high-level theory rarely even think of rejecting or disconfirming those models; the very idea of trying to disprove practice theory or materiality sounds absurd. There is nothing wrong with high-level theory per se, particularly for those interested in philosophical ideas about the human condition. But if one is going to build rigorous archaeological arguments that have a chance of being wrong, then high-level theory is useful mainly for providing a context for the middle-range theory that does the explanatory work.

Causal Mechanisms

In the decades since the demise of the covering-law model for social science explanation, a new approach to explanation has flourished in the social sciences: causal mechanisms. In the words of philosopher of science Mario Bunge (2004:182), "to explain a fact is to exhibit the mechanism(s) that makes the system tick." The analysis of causal mechanisms is closely linked to middle-range theory, both conceptually and in practice. There is a large and expanding literature on causal mechanisms in sociology, political science, social history, and other fields, but archaeologists almost never cite this material. Yet if we want to make convincing arguments about social conditions and changes in the past, it is well worth our while to explore the ways social historians such as Charles Tilly (2008) analyze and explain social continuity and change by way of causal mechanisms.

Discussion

Perhaps the most basic stipulation of good social-science argumentation is expressed in the epigraph: You have to be able to tell when you are wrong. This can be achieved by formal hypothesis testing or by more informal methods that evaluate alternative explanations. Although some archaeologists do not approve of this emphasis on testing (Johnson 2010:223), it is hard to see how we can achieve rigorous results and build a solid body of knowledge without it. In this paper, I advocate that archaeologists maintain an explicit consideration of the structure of their arguments, while avoiding pitfalls such as ad-hoc analogies and empty citations. The methods of strong inference and natural

experiments hold promise for archaeology, particularly when used together with what social scientists (but not Lewis Binford) call middle-range theory.

Acknowledgments. I would like to thank the students in my graduate seminar on Theory for Archaeology for helpful discussions about the issues discussed in this paper. Of the works discussed here, they found the papers by Abend (2008) and Wylie (1985) to be the most useful and insightful. I appreciate the feedback on an earlier draft from several colleagues, including Michelle Hegmon, Keigh Kintigh, and, especially, Chris Morehart.

References Cited

Abbott, Andrew
2004 *Methods of Discovery: Heuristics for the Social Sciences*. Norton, New York.

Abend, Gabriel
2008 The Meaning of "Theory." *Sociological Theory* 26:173–199.

Binford, Lewis R.
1967 Smudge Pits and Hide Smoking: The Use of Analogy in Archaeological Reasoning. *American Antiquity* 32(1):1–12.
1981 *Bones: Ancient Men and Modern Myths*. Academic Press, New York.

Booth, Wayne C., Gregory G. Colomb, and Joseph M. Williams
2008 *The Craft of Research*. 3rd ed. University of Chicago Press, Chicago.

Bunge, Mario
2004 How Does It Work? The Search for Explanatory Mechanisms. *Philosophy of the Social Sciences* 34(2):182–210.

Chamberlin, Thomas C.
1965 The Method of Multiple Working Hypotheses. *Science* 148:754–759.

Copi, Irving M.
1982 *Introduction to Logic*. 6th ed. Macmillan, New York.

Davis, Rowland H.
2006 Strong Inference: Rationale or inspiration? *Perspectives in Biology and Medicine* 49(2):238–250.

Diamond, Jared, and James A. Robinson (editors)
2010 *Natural Experiments of History*. Harvard University Press, Cambridge.

Fogelin, Lars
2007 Inference to the Best Explanation: A Common and Effective form of Archaeological Reasoning. *American Antiquity* 72:603–625.

Gerring, John
2007 *Case Study Research: Principles and Practices*. Cambridge University Press, New York.
2012 *Social Science Methodology: A Unified Framework*. 2nd ed. Cambridge University Press, New York.

Haber, Stephen
1999 Anything Goes: Mexico's "New" Cultural History. *Hispanic American Historical Review* 79:309–330.

Harzing, Anne-Wil
2002 Are Our Referencing Errors Undermining our Scholarship and Credibility? The Case of Expatriate Failure Rates. *Journal of Organizational Behavior* 23:127–148.

Johnson, Matthew
2010 *Archaeological Theory: An Introduction*. 2nd ed. Blackwell, Oxford.

Kerr, Norbert L.
1998 HARKing: Hypothesizing after the Results Are Known. *Personality and Social Psychology Review* 2(3):196–217.

Kintigh, Keith W., Jeffrey Altschul, Mary Beaudry, Robert Drennan, Ann Kinzig, Timothy Kohler, W. Frederick Limp, Herbert Maschner, William Michener, Timothy Pauketat, Peter Peregrine, Jeremy Sabloff, Tony Wilkinson, Henry Wright, and Melinda Zeder
2014 Grand Challenges for Archaeology. *American Antiquity* 79(1):5–24.

Merton, Robert K.
1968 *Social Theory and Social Structure*. 3rd ed. Free Press, New York.

Platt, John R.
1964 Strong Inference. *Science* 146:347–353.

Raab, L. Mark, and Albert C. Goodyear
1984 Middle-Range Theory in Archaeology: A Critical Review of Origins and Applications. *American Antiquity* 49:255–268.

Ragin, Charles C., and Lisa M. Amoroso
2011 *Constructing Social Research: The Unity and Diversity of Method*. 2nd ed. Sage, Thousand Oaks, California.

Smith, Michael E.
2011 Empirical Urban Theory for Archaeologists. *Journal of Archaeological Method and Theory* 18:167–192.
2016 *At Home with the Aztecs: An Archaeologist Uncovers Their Domestic Life*. Taylor and Francis, New York, in press.

Tilly, Charles
2008 *Explaining Social Processes*. Paradigm Publishers, Boulder, Colorado.

Wylie, Alison
1985 The Reaction against Analogy. *Advances in Archaeological Method and Theory* 8:63–111.
2000 Questions of Evidence, Legitimacy, and the (Dis)unity of Science. *American Antiquity* 65:227–237.