

Bootstrapping in Un-Natural Sciences:

Archaeological Theory Testing¹

Alison Wylie

University of Western Ontario

1. Introduction

Glymour's bootstrapping account of confirmation is meant to show how it is that evidence can bear on a theory in a discriminating, non-circular way even when that theory is used to establish the inferential link between evidence and a test hypothesis. Evidence confirms a theory on his account if, "using the theory, we can deduce from the evidence an instance of the hypothesis i.e., an hypothesis comprising or instantiating the test theory, and the deduction is such that it does not guarantee that we would have gotten an instance of the hypothesis regardless of what the evidence might have been." (1980, p. 127). Glymour goes on to argue that this strategy of inference should appear most explicitly in the developing and "un-natural" (social) sciences where novel theories are being formulated or applied to new domains (1980, p. 172). Here, he says, there will be little in the way of developed "substantive principles about the bearing of evidence" (1980, p. 291) to obscure the essential bootstrapping structure of confirming arguments.

On the face of it, however, the specifics of Glymour's model would seem to make it quite implausible that bootstrapping should be especially prevalent and explicit in the developing sciences. It is, for example, an important structural requirement of his account that test theories provide a determinate computation of values for all relevant variables and yet, as van Fraassen has argued (1983, p. 32-33), developing sciences can rarely meet such a condition. This is particularly true of "un-natural" sciences; not only are researchers in these fields often unable to specify relations among variables closely enough to allow calculation of their values from one another (the situation which, I take it, concerns van Fraassen), they are even frequently uncertain what range of variables must be taken into account. It would seem, then, that insofar as bootstrap strategies are employed in these contexts, they will necessarily diverge from Glymour's model in a number of significant respects.

I propose to examine an influential example of archaeological testing that does conform to Glymour's model in intent and, in broad outline at

PSA 1986, Volume 1, pp. 314-321

Copyright © 1986 by the Philosophy of Science Association

least, in practice, with the aim of specifying how bootstrap strategies function when theory is not just "becoming more testable" (van Fraassen 1983, p. 33) but is in the initial stages of development (or is undergoing extensive reformulation). There are three inter-dependent respects, I shall argue, in which bootstrap practice departs from the ideal in such testing situations: 1) testing is not strictly theory contained, 2) the theory-mediated inference from evidence to test hypothesis is not exclusively deductive and, 3) structural considerations do not displace or take precedence over substantive considerations. What emerges by way of a constructive thesis is that bootstrapping in developing and exploratory sciences is as much a process of theory construction as it is of theory testing.

2. Archaeological Testing

Archaeology is a paradigmatically un-natural and crisis ridden field. In the late 1960s a group identified as the "New Archaeologists" rejected the theoretical commitments and methodology of "traditional" practice and promoted a "scientific" research program. It was modelled on what Glymour would no doubt call a "fantasy image of physics" (1980, p. 292); archaeologists were to abandon reliance on "inductive" methods of interpretation and institute a program of strict (positivistic) testing through the confrontation of hypotheses with the archaeological "facts". Difficulties encountered in implementing this program have made it clear, however, that archaeological data constitute test evidence only under interpretation and that interpretation is very often mediated by auxiliary hypotheses drawn from the same general theory of cultural phenomena as underlies (i.e., incorporates as an element or entails as an instance) the more specific hypotheses about the cultural past that archaeologists are generally concerned to test. This is just to take Glymour's point that testing is a three-place relation in which confirming arguments move from evidence to test hypothesis. It is also to raise exactly the kind of problem about circularity that Glymour claims bootstrap testing can circumvent when he insists that bootstrapping "should be relevant wherever arguments about the possibility or impossibility of knowing something turn on questions of alleged circularity." (1980, p. 376).

The response of the New Archaeologists to this problem is strikingly like Glymour's. They have been resolute in denying that such theory-dependence is necessarily vicious in its circularity; they insist that archaeological data can provide discriminating evidence for or against a test hypothesis so long as the mediating theory establishes a determinate relationship between the values of measurable (material) and hypothetical (cultural/behavioral) variables. The problem of building a body of theory capable of securing the "deduction" of hypotheses from the evidence has taken center stage in the last few years. But even before this, New Archaeologists routinely took care to develop plausible theoretical arguments about the evidential import of test data and to establish that these arguments would not arbitrarily ensure confirmation whatever the empirical results of inquiry.

This commitment to institute what looks very much like a bootstrap testing methodology is closely aligned with rejection of "normativist" assumptions about the cultural subject associated with traditional research. Rather than seeing human behavior (and its material remains)

structured primarily by a repertoire of culture-specific "ideals" (norms) or conventions, the New Archaeologists promoted a "systemic" and essentially materialist theory of culture. Culture is most plausibly conceived, they argued, as a complex system of mutually conditioning elements (including material, social, and technological as well as ideational components) which is shaped primarily in response to material conditions of life; culture is, fundamentally, "man's (sic) exosomatic means of adaptation" (Binford 1972, p. 136) or, alternately, "a material-based organization of behavior" (Binford 1972, p. 8). Although this theoretical framework was initially endorsed as a necessary presupposition of research, its use as a source of explanatory models of particular past contexts (the models which New Archaeologists were to test archaeologically) was itself treated as a test, usually a vindication, of its general, substantive claims about the nature of cultural systems. This is clearly one concern that informed the design of a series of now classic excavations undertaken by New Archaeologists in the American Southwest. I want to consider the research carried out by Bill Longacre and Jim Hill at two 12th-13th century pueblos, Carter Ranch and Broken K, in the Hay Hollow valley.

The theoretical problem of interest in this area was to explain the widespread phenomena of population decrease and aggregation that took place throughout the Southwest immediately before and during the time the Carter Ranch and Broken K Pueblos were occupied (circa AD 1100 to 1280) and that resulted, after AD 1300, in abandonment of most of the region. The standard hypotheses about such dramatic collapse were unsupported; there was no evidence of invasion or violent internal conflict (such as would require aggregation in defensible villages) or of extensive disease, and there was no indication of catastrophic change in the environment on the scale of the cultural events to be explained (e.g., region-wide resource depletion or extensive drought).

Given their materialist commitments the New Archaeologists were inclined to entertain some version of the last hypothesis, and their "systemic" conception of culture suggested that a relatively less dramatic change in environmental conditions than had been envisioned might well have served as a trigger, setting off a sequence of local and restricted adjustments the cumulative effect of which might be the large scale transformation of pueblo culture documented archaeologically. They thus attributed greater significance than "traditional" researchers had done to paleo-environmental evidence of a region-wide shift in the pattern, but not overall amount, of rainfall; gentle, dispersed winter rainfall gave way to torrential summer storms of a sort that would have increased erosion and diminished the effective surface moisture. While this would not have compromised agricultural production across the whole region, it would have begun to restrict maize production in the more marginal (upland) areas after AD 1100 which, in combination with population pressure, could have quickly created local shortfalls. The hypothesis that emerged in the research of the 1960's and 1970's was that one of the few viable responses open to those who resisted return to a fully mobile collecting economy would have been the development (or, increased exploitation) of social mechanisms for pooling regional resources, including exchange, increased inter-site co-operation and, eventually, aggregation. The dramatic aggregation and decline of the population in succeeding generations would then be explicable as a culturally mediated response to gradual and relatively undramatic

changes in the environment, consistent with the encompassing eco-system model of culture.

This hypothesis and more generally, the eco-system theory it instantiates, directed attention to a number of variables that had not previously been analysed or reconstructed in any detail: finegrained shifts in patterns of resource exploitation (which might reflect environmental pressure), internal intra-site and intra-assemblage variability (which might indicate something of the social structure and level of integration of prehistoric pueblo communities), and intra-regional trade networks (suggesting a system of redistribution that might have buffered those in areas of shortfall). In connection with the first, Hill was able to establish that the occupants of Carter Ranch and Broken K, the two largest and latest sites in the Hay Hollow Valley, were under increasing resource pressure during the period immediately before abandonment; the faunal data and plant remains showed a consistent decrease in dependence on cultigens and large game and a corresponding increase in dependence on wild plants and small game. Hill and Longacre were among the first to attempt to investigate the second factor (the third is currently a dominant research concern) and their results here are most striking. Longacre's earlier study at Carter Ranch Pueblo established a highly significant statistical association of ceramics painted with distinctive clusters of design elements with three separate sectors of the pueblo. He argued that this could not be accounted for functionally or temporally (i.e., the style differences do not correspond to activity areas or to different periods of occupation) but that it might plausibly be explained as a function of social differentiation within the pueblo community. His specific hypothesis was that by AD 1100 the matrilineal residence system and perhaps the associated matrilineal system of descent typical of contact period pueblos had already been established (some 100-300 years earlier than postulated by ethno-historic reconstruction; Hill 1970, p. 74) but that nothing like the level of social integration typical of these later pueblos had been achieved; formerly autonomous and dispersed lineage units eco-existed in single village settlements but retained their social distinctness. The significance of this, if it could be established, is that it strongly suggests that aggregation was, indeed, a response to environmental pressure, not a function of independent shifts in dominant social norms.

Because neither the orienting theory nor the test hypothesis incorporates any established general principles about the relationship between the material and social variables in question, the linking arguments Hill and Longacre use to establish their hypothesis of internal social differentiation depends heavily on ethnographic analogy. It runs as follows: if, as in modern pueblos, women were the primary producers of ceramics and passed on design styles generationally (i.e., if they learned design styles primarily from their mothers), then a localization of ceramic design would occur only if kinswomen lived in cross-generationally stable residential groups, as under a matrilineal residence system. Hill replicated Longacre's results in analysis of the ceramic data from Broken K Pueblo, and then undertook to test for corroborating patterns of stylistic differentiation and distribution in other classes of artifacts typically associated with female activities. The result was striking confirmation that pueblos of the period were comprised of socially distinct residential units. Given the size of

these sub-components they seemed, specifically, an amalgam of the village and "homestead" units that had previously been dispersed throughout the region, now co-existing next to the most stable supply of water in the valley in a final effort to survive in the area as sedentary agriculturalists. The outcome of the research is, then, not only confirmation but also a specification of the details of an explanatory hypothesis that was initially just a sketch along lines suggested by the encompassing theory. The New Archaeologists have been quick to claim that the empirical credibility of the theory, which gains content and specificity, is thereby significantly increased.

3. Constructive Bootstrapping

At all levels of analysis it is clear that Hill and Longacre construct confirming arguments that move from evidence to test hypothesis, not the reverse (this is unavoidable inasmuch as these arguments are partly constitutive of the hypothesis), and that depend explicitly on linking assumptions about the nature of the evidence in question and its relation to the conditions postulated by the test hypothesis. To this extent, their arguments fit Glymour's model. It is also clear, however, that bootstrapping in this context is not a matter of using the resources of a single subject-specific theory to establish tests of its own empirical adequacy. Not only is the theory in question incomplete, the range of conditions responsible for the production of an archaeological record is so great that even if it were complete and comprehensive, it could not be expected to specify relationships between the full range of variables that archaeological linking arguments must consider. In the cases discussed, it was crucial to reconstruct certain non-cultural variables--environmental conditions and patterns of resource exploitation--which required appeal to independent bodies of scientific theory (primarily biology and ecology). Absolute dating of all kinds, and reconstructions of prehistoric technology and subsistence regimes routinely appeal to collateral theory of this sort. Even where the variables in question are specifically cultural, the applicable theories, usually drawn from anthropology, are notoriously incomplete in the areas of particular interest to archaeologists; they may identify all the relevant socio-cultural variables and their interrelations, but typically do not specify relations between them and the material variables accessible to archaeologists. In assessing the evidential import of their data archaeologists must routinely appeal directly to background knowledge and ethnographic instances, as Hill and Longacre did, for an understanding of the socio-cultural conditions that could have produced the record; this is information that would be subsumed (as supporting evidence) by general linking principles and incorporated into the theory if it were fully developed. As things stand, however, bootstrap testing in a discipline like archaeology is not and perhaps could not ever be theory-contained in the manner required by Glymour's model.

This open-endedness is at once a source of difficulty and strength. The difficulty is that arguments concerning the evidential import of archaeological data are bound to be inconclusive. Glymour's ideal of confirmation by deductive linking arguments may perhaps be approximated in archaeology when the hypothesis under test is exclusively concerned with the bio-physical conditions responsible for the archaeological record or with human behaviors that are completely determined by such

conditions. Inferences concerning the paleo-environment and some aspects of subsistence-related behavior may be examples of this in the Hill, Longacre studies. Although these kinds of inferences are crucially important when they can be made they cover only a very small range of the phenomena archaeologists typically seek to reconstruct and explain. In the vast majority of cases, especially those where exclusively social, cultural variables are concerned, archaeologists rely on some form of abductive inference which is typically analogical. In this, archaeological arguments of confirmation consistently depart from Glymour's deductive ideal.

The strength of such arguments, which New Archaeologists are increasingly concerned to exploit, is that when they draw on resources external to the theory under test, they are in a position to set up a system of internal constraints between different lines of supporting evidence. This promises not only a check on the accuracy of specific linking assumptions but, when conciliation emerges, a dramatic improvement on the constructive support that any one type of test evidence could provide considered on its own (Meehl discusses this in consideration of "consistency tests", 1983). When, for example, inferences mediated by sources as diverse as bio-ecological theory, pueblo ethnography, and theories about cultural evolution and adaptation all yield the hypothesis that pueblo aggregation was a response to environmental stress, Hill and Longacre's test data provides particularly strong confirmation; it is highly implausible that such conciliation could be an artifact of theoretical expectation.

It is important to recognize, however, that this strength derives from the convergence of substantive considerations of exactly the sort that Glymour insists are secondary and incapable of accounting for "the fine points of the distribution of praise and blame among hypotheses" (1980, p. 375). Faced with a lack of developed theoretical understanding in the relevant areas, Hill and Longacre do not resort to structural considerations but rather, to a more tentative, ad hoc, and particularistic form of substantive consideration. In their confirming arguments, analogical inference serves very clearly as a means of importing empirical information about the nature of their evidence and how it might have been produced. It is hard to see how this could be otherwise. How could a theory be developed that specifies the relations holding among component variables in the absence of substantive knowledge of the subject domain in question? It would seem that the structure of a theory and of inferences that bootstrap confirm are unavoidably parasitic on substantive considerations of content. This suggests that Glymour's emphasis on the primacy of structural considerations is misplaced, even (or especially) for unnatural sciences at early stages of development.

It is also important to see that the use of analogical arguments to import substantive considerations has a constructive aspect that Glymour overlooks. For Longacre, the ethnographic data on pueblo ceramic production served, in the first instance, as the source of fragmentary insights about links that might hold between his archaeological data and the social organization of prehistoric pueblo communities. In order to use a wider range of artifact types to test this hypothesis about social organization, Hill generalized on this insight; his linking argument appeals directly to the hypothesis that stylistic similarity (at the

level of subconscious preference for the smallest units of design) is an index of intensity of social interaction (Plog 1980). The discovery that this principle anticipates and makes sense of patterning in a much wider range of artifact classes than originally considered not only confirms the test hypothesis but also at least "reduces the uncertainty" of the linking hypothesis itself (to paraphrase Horwich, 1983). The immediate import of this is that it suggests the existence of "ancestral" relations among hypotheses with regard to evidence (van Fraassen 1983) by which, contrary to Glymour's model, confirmation extends, at least weakly, to the conjuncts of a successful test hypothesis. More generally, it suggests that, in Hill and Longacre's hands, bootstrap confirmation is a process not just of testing an hypothesis that instantiates their developing theory, but of building into this theory the resources it needs to test such an hypothesis and, thus, to raise itself confirmationally by its own bootstraps. Glymour's focus on structure obscures precisely the features of this process--the open-endedness, the reliance on analogy, and the centrality of substantive considerations--that are essential to its constructive function.

My conclusion is, then, that bootstrap confirmation in developing sciences is not only a reflexive, probative strategy for evaluating novel theories, but also, and necessarily, a process of using empirical and theoretical knowledge established in a variety of contexts to build and refine such theories. The judgments researchers render concerning the bearing of evidence are, therefore, irreducibly a function of the background information that they have available (or recognise as relevant). As such, they constitute not simply an assessment of the credibility of discrete components of an encompassing theory but also an evaluation of how a given theory may most fruitfully be developed.

Notes

¹The research for this paper was partially supported by a Postdoctoral Fellowship at the University of Calgary. Earlier versions of the paper were read to the Department of Philosophy at the University of Western Ontario (January 1985) and at the Canadian Philosophical Association Meetings (May 1985); I thank those present on these occasions for extremely valuable discussion. In particular, I thank Kathleen Okruhlik for her thoughtful and detailed commentary at the CPA.

References

- Binford, L. R. (1972). An Archaeological Perspective. New York: Seminar Press.
- Earman, John. (ed.). (1983). Testing Scientific Theories. (Minnesota Studies in the Philosophy of Science, Volume X.) Minneapolis: University of Minnesota Press.
- Glymour, Clark. (1980). Theory and Evidence. Princeton N.J.: Princeton University Press.
- (1983). "On Testing and Evidence." In Earman (1983). Pages 3-26.
- Hill, J. N. (1966). "A Prehistoric Community in Eastern Arizona." Southwestern Journal of Anthropology 22: 9-30.
- (1968). "Broken K. Pueblo: Patterns of Form and Function." In New Perspectives in Archaeology. Edited by L.R. Binford and S.R. Binford. Chicago: Aldine. Pages 103-142.
- (1970). Broken K Pueblo: Prehistoric Social Organization in the American Southwest. (Anthropological Papers of the University of Arizona # 18.) Tucson: University of Arizona Press.
- Horwich, Paul. (1983). "Explanations of Irrelevance." In Earman (1983). Pages 55-66.
- Longacre, W. A. (1966). "Changing Patterns of Social Integration: A Prehistoric Example from the American Southwest." American Anthropologist 68: 94-102.
- (ed.). (1970). Reconstructing Prehistoric Pueblo Social Organization. Albuquerque: University of New Mexico Press.
- Meehl, Paul E. (1983). "Consistency Tests in Estimating the Completeness of the Fossil Record." In Earman (1983). Pages 413-473.
- Flog, S. (1978). "Social Interaction and Stylistic Similarity: A Reanalysis." In Advances in Archaeological Method and Theory, Volume 1. Edited by Michael B. Schiffer. New York: Academic Press. Pages 143-181.
- (1980). Stylistic Variation in Prehistoric Ceramics: Design Analysis in the American Southwest. New York: Cambridge University Press.
- van Fraassen, Bas C. (1983). "Theory Comparison and Relevant Evidence." In Earman (1983). Pages 27-42.