Sunk-Cost Effects and Vulnerability to Collapse in Ancient Societies

MARCIO A. JANSSEN, TIMOTHY A. KOHLER, AND MARTEN SCHEFFER
Center for the Study of Institutions, Population, and Environmental Change, Indiana University, 408 N. Indiana Ave., Bloomington, Ind. 47408-3799, U.S.A. (maajans@indiana.edu)/Department of Anthropology, Washington State University, P.O. Box 644910, Pullman, Wash. 99164-4910, U.S.A. (tako@wsu.edu)/Department of Aquatic Ecology and Water Quality Management, Wageningen University, P.O. Box 8080, NL-6700 DD Wageningen, The Netherlands (marten.scheffer@aqec.wkao.wau.nl).

Judging by the variety of explanations proffered in the literature, societies apparently collapse for a variety of reasons. In an influential review, Tainter (1988:39–50) reports that published explanations for the famous collapses of the Classic Maya, the western Roman Empire, and many other less famous episodes such as the demise of the Chimu of Peru or the Chacoan system in the U.S. Southwest tend to fall into 11 major categories, with resource depletion or deterioration being one of the most commonly adduced causes (see, for example, Hodell, Curtis, and Brenner 1995; Weiss and Bradley 2001). Another common explanation is insufficient response to circumstances, or “failure to adapt.” Tainter [1988:50, emphasis added] rejects both such explanations. Resource-degradation explanations raise the question why societies sit by and watch the encroaching weakness without taking corrective actions. . . . As it becomes apparent to the members or administrators of a complex society that a resource base is deteriorating, it seems most reasonable to assume that some rational steps are taken towards a resolution. . . . If a society cannot deal with resource depletion (which all societies are to some extent designed to do) then the truly interesting questions revolve around the society, not the resource. What structural, political, or economic factors in a society prevented an appropriate response?

As to the possibility that societies fail because they are inherently fragile or static or incapable of shifting directions, Tainter (1988:54–61, 89) considers this not so much an explanation as something that, if true in particular cases, must be explained.

SUNK-COST EFFECTS: INDIVIDUALS AND GROUPS

In this report we seek to unite these two explanations in a model that suggests why and under what conditions societies faced with resource degradation might “fail to adapt.” We are not peddling a new universal theory for societal collapse; we do hope to insert into the anthropological conversation on collapse a mechanism—little noticed to date—making some societies more vulnerable to collapse under certain conditions. Our model, which we illustrate with a simple mathematical characterization, is based on a well-documented systematic deviation from rational decision making known as the “sunk-cost effect” [Arkes and Ayton 1999]. Rational-choice theory tells us that one’s choice between options should be influenced not by prior investment but only by the expected future costs and benefits of those options. Of course, prior investment may affect the knowledge and experience of the decision maker, but such effects can be included in the rational-choice theory explanation of decision making. Numerous studies [Arkes and Blumer 1985, Arkes and Ayton 1999, Teger 1980] nevertheless demonstrate that humans do consider prior investment in deciding what course of action to take.

Most sunk-cost research is focused on individual decision making. Explanations of observed escalation of commitment are self-justification, not admitting that past decisions were incorrect, and framing effects (whether the problem is framed in positive or negative outcomes) [Staw 1997]. One might expect that such irrational behavior could be corrected in groups. Groups, however, are actually more prone than individuals to succumb to the sunk-cost effect [Whyte 1993]. Nor is this effect limited to Western society; conformist (or frequency-dependent) transmission is a leading candidate for explaining behavior (including self-sacrificial cooperation in large groups) that is otherwise difficult to account for [Boyd and Richerson 1985:204–40]. Indeed, a typical goal for political decisions in small-scale societies is consensus [Boehm 1996]. Once members of a group reach consensus, the easiest way to maintain it is to stay committed to the group’s decision [Janis 1972]. Thus, even when the group is faced with negative results, members may not suggest abandoning an earlier course of action, since this might break the existing unanimity.

© 2003 by The Wenner-Gren Foundation for Anthropological Research. All rights reserved 0011-3304/2003/4405-0051$1.00

1. We thank Marty Andries, Jared Diamond, Katja Hogendoorn, Garry Peterson, Henry Wright, and four anonymous reviewers for their useful suggestions. We began this collaboration following acquaintances made at a Santa Fe Institute–sponsored Long-Term Human Dynamics Working Group meeting. Janssen acknowledges support from NSF SBR9531018 and Kohler from NSF BCS-0119981 and many useful conversations about the northern Southwest with W. D. Lipe and M. D. Varien. [Supplementary material appears in the electronic edition of this issue on the journal’s web page [http://www.journals.uchicago.edu/CA/home.html].]
A Simple Model for Resource Dynamics Under Harvest

There is, then, abundant empirical evidence that humans become increasingly unlikely to abandon a (failing) course of action to the extent that they have more investment in it and that this effect tends to be amplified by group processes. We propose that this dynamic may lead to the postponement of small adaptive adjustments until more dramatic changes become necessary. To see how this effect might lead to collapse of settlements, suppose that the dynamics of the local renewable resources \( R \) used by a settlement of humans \( H \) can be described by a classic model of a logistically regrowing resource exploited by a consumer such as

\[
\frac{dR}{dt} = gR \left( 1 - \frac{R}{K} \right) - c_{\text{max}} \frac{R}{R + h_i} H + i,
\]

where \( g \) is the maximum growth rate, \( K \) the maximum level of the resource, \( c_{\text{max}} \) the maximum per capita human resource consumption, and \( h_i \) the resource level needed to reach 50% of the maximum consumption rate. Because networks of trade may buffer local resource decline, we add resupply of local resources \( i \) from neighboring areas, which may prevent irreversible extinction of local resources [Scheffer and de Boer 1995]. If consumers in such models are efficient exploiters \( h_i \) small relative to \( K \), the equilibrium resource level will be a sigmoidal function of the population size of humans [fig. 1] as described for various ecosystems by Scheffer et al. [2001]. In our case it implies that over a range of human population densities two alternative stable states are possible: one with a relatively high resource level on the upper branch and an alternative overexploited state with a low resource level on the lower branch. In between are unstable equilibria that mark the critical resource level below which recovery is possible only when population size is reduced to a level smaller than \( F_1 \).

We will not consider human population dynamics in detail but simply assume that the population in the settlement tends to grow if resources are abundant and decrease if the resource level falls below a certain critical limit \( R_c \) when individuals quit the settlement in search of better opportunities. We may now analyze the dynamics graphically through a "slow-fast approach" [Rinaldi and Scheffer 2000], if we consider the dynamics of human settlement size \( H \) to be slow relative to the dynamics of the resource. This appears to be a reasonable assumption for many agricultural societies in which production may vary greatly from year to year as it does in the pre-Hispanic Pueblo dry-farming regimes of southwestern North America [Van West 1993] from which our examples will come.

Plotting the critical resource level \( R_c \), below which humans quit a place and the resource equilibrium curve, we can explore the expected effect of investment in fixed structures (for example, temples, other public structures, and housing) on the dynamics of settlements (fig. 3). The sunk-cost effect implies that the critical resource level tolerated before abandoning the site should decrease if more has been invested in fixed structures. Thus, if not much has been invested in local settlements, this level will tend to be high [fig. 2, a], resulting in a stable equilibrium at the intersection of the two zero-growth isoclines. If sunk-cost effects cause people to leave only at a lower resource level, this equilibrium shifts to higher population densities [fig. 2, b]. At even stronger sunk-cost effects the interception is in the unstable part of the resource curve (fig. 2, c). In this case the settlement will grow until point \( F_2 \) is reached and the resource crashes, causing the abandonment of the settlement. Eventually the resources will recover and new settlements may be set up in the same area or nearby, resulting in a cyclic development. Inputs from outside the region, for example, by trade, will shift the curves in these figures to the right. The system may reach higher population levels because of such external inputs, but the qualitative impact of the sunk-cost effect remains the same.

One may alternatively interpret figure 2, c, as a system that becomes vulnerable to adverse events because of increasing settlement size. Disturbances such as floods, droughts, or conflict with neighboring groups may push the system over the edge, leading to a collapse. Stochastic events that reduce resource abundance (e.g., pests, fires, etc.)
Fig. 2. The expected effect of investment in fixed structures on the dynamics of settlement. The critical resource level ($R_c$) below which the human population in a settlement declines defines the zero-growth isocline ($H = 0$) of the settlement. Assuming human dynamics to be slow relative to those of resources, we analyze the stability of the equilibria at intersections with the resource isocline ($R = 0$). a, no sunk-cost effects. b, moderate sunk-cost effects. c, stronger sunk-cost effects.

Droughts may have little effect in small settlements but in large settlements can easily bring the system across the border of the attraction basin of the overexploited state, resulting in a crash. Thus the model predicts that sunk-cost effects can lead to growth of settlements to a point where they are about to overexploit their resources. At this point resilience (the basin of attraction) becomes very small and adverse stochastic events will tend to induce collapse.

Evidently, the actual value of the parameters in such abstract models will be difficult to assess in practice, but analyses with various minimal models of this form [not shown] indicate that the behavior occurs over a wide range of parameter values and alternative models. For instance, our analysis of an alternative model in which human migration dynamics were included explicitly and the critical point for leaving a settlement depended on realized consumption ($C$) rather than resource level ($R$) yielded the same qualitative results. We also obtained similar results from a more elaborate model including the economics of societies and the dynamics of investment in settlement structures [Janssen and Scheffer 2003]. It thus appears that the prediction is quite robust against details of the models used: societies that invest heavily in structures, monuments, or even equipment and facilities for very specific extractive activities become liable to collapse through sunk-cost effects from resource overexploitation.

Identifying sunk-cost effects in the archaeological record

The predictions of this hypothesis may be examined against the archaeological record. Here we briefly present evidence for this effect from pre-Hispanic Pueblo (Anasazi) populations, which constructed some of the largest nonearthen structures built in the United States before the Chicago skyscrapers of the 1880s [e.g., Pueblo Bonito, Chaco Canyon, New Mexico [Lekson 1984]]. Our general strategy will be to contrast timing of new construction investments in locations with differing prior construction investments, with the expectation that large prior investments will [improperly, from a rational economic perspective] induce further investment even in periods of scarcity.

Despite their accomplishments these are relatively small-scale societies, and it might be argued that “collapse” is not an appropriate term in the cases to be discussed. We agree that it is not always clear, in the Southwest, what should be considered a collapse. Movement of habitations and the communities they constituted was routine in the northern Southwest, and therefore what looks like collapse locally may appear from a regional perspective to be a resilient adaptation [Kohler and Matthews 1988, Nelson and Hegmon 2001]. Less commonly, but typically in the three cases we examine, groups of large villages were abandoned rather suddenly and not replaced locally with systems of similar complexity. It is perhaps a little easier to see these as episodes of collapse even if, in some cases, systems of similar com-
plexity emerged elsewhere in the region. More generally, though, we agree with Tainter (1988:4-5) that collapse is not a process restricted to complex societies, although they certainly provide us with particularly dramatic examples. The pre-Hispanic Puebloan cases are in fact especially suitable for detecting even rather subtle sunk-cost effects, because tree-ring analysis can frequently provide estimates of potential annual agricultural production [production on which these societies were extremely dependent] and high-resolution dating of construction activity in both small and large settlements.

In the U.S. Southwest, pre-Hispanic agricultural Puebloan peoples prior to A.D. 1300 constructed many hamlets of just a few households and larger villages2 with evidence of large and labor-costly "public" structures such as great kivas, oversized pithouses, D-shaped and other multiwalled structures, reservoirs, enclosing walls, and so forth, lacking at smaller habitations. The sunk-cost hypothesis predicts that people will continue to invest in construction at large settlements even into periods of scarcity, whereas construction at small settlements should be more confined to periods of relative abundance, given the slighter investment in local facilities. This should be true even if these hamlets and villages are part of the same community [in which case the hamlet dweller may be relocating to the village in bad times] so long as the hamlet dweller can be considered to have more investment in the community center [village] than in an outlying habitation.

The area near the town of Dolores in southwestern Colorado (fig. 3) was intensively studied by the Dolores Archaeological Program from 1978 to 1985 [Breternitz, Robinson, and Gross 1986]. The area is both high and near the local northeastern boundary of the possible zone for maize agriculture [Petersen 1986]. Most of the numerous Puebloan sites in this area were occupied between A.D. 650 and 900. As elsewhere in the Puebloan world between A.D. 700 and 1300, two obvious classes of residential sites existed: hamlets of one or a few households and villages of a dozen or more. Eight large villages are evident in the Dolores River valley in the mid-800s, with the largest, McPhee, composed of nearly 200 households. During the height of the Dolores-area occupation in the mid-to-late 800s there is evidence of depletion of wood resources, piñon seeds, and [less clearly] large game, compensated for agricultural intensification [reviewed by Kohler 1992].

The Dolores area lost population rapidly after the 880s, displacing several thousand individuals [Kane 1986:370]. Given the heavy reliance on agriculture, a series of locally dry and regionally cold years in the late 800s [Petersen 1986; Salzer 2000:132-71] contributed, at least, to this depopulation. The small number of people who stayed, primarily at Grass Mesa Village during the "Grass Mesa Subphase" [≈ A.D. 880-910; Lipe, Morris, and Kohler 1988], both increased their mobility and underwent considerable organizational simplification. In figure 4, we graph construction timber procurement events [Schlanger and Wilshusen 1993] in the Dolores area against a standardized proxy measure for agricultural productivity. As predicted by the sunk-cost effect, construction in hamlets is restricted primarily to years in which productivity is at or above the long-term mean, whereas construction in villages persists under highly variable [and even poor] conditions. Following the collapse of these Pueblo I period villages, the Dolores area was never reoccupied in force by Puebloan farmers.

It is unlikely that the pattern we see in Dolores is due to differing microenvironments for villages and hamlets that would make the stresses more severe in hamlets. Habitations between A.D. 600 and 920, both hamlets and villages, tend to be located so as to maximize the amount of good-quality agricultural land within their 1-km catchments and to minimize the amount of poor-quality agricultural lands [Orcutt 1987:661]. Further, an analysis of agricultural costs across the project area demonstrates that hamlets that tend to grow into villages are neither exceptionally favorable nor unfavorable locations for agriculture. The result of aggregation at such places, therefore, is to increase average per household distances to fields considerably [Kohler et al. 1986]. Hence, it is probably safe to assert that the decision to remain in villages toward the end of the occupation could not have been motivated by narrow calculations of minimizing agricultural costs. We suggest that additional investment in such sites under these conditions is an example of sunk-cost effects.

A somewhat similar case comes from nearby Sand Canyon Locality west of Cortez, Colorado [fig. 3], intensively studied by the Crow Canyon Archaeological Center over the past 15 years [Lipe 1992]. Here the main occu-

---

2. We restrict our analysis to villages before A.D. 1300 because the often considerably larger post-A.D. 1300 pre-Hispanic Puebloan towns, at least along the Rio Grande, tend to be more permanent [arguably because of a different economic organization [Kohler, Van Pelt, and Yap 2000]], are not well enough dated to support this kind of analysis, and often have rather unclear demographic histories.
Fig. 4. Timber procurement events (TPEs) in the Dolores area (top) and the Sand Canyon Locality (bottom), graphed against proxies for agricultural production derived from tree rings and smoothed to the mean of the current year plus the previous two. Dot, a TPE at a village; circle, a TPE at a hamlet. Lines connect TPEs from the same village; dotted lines connect TPEs from the same hamlet. In hamlets, timber is harvested only in relatively productive years, whereas in villages construction apparently continued in periods of poor agricultural production. Lines connecting TPEs from the same sites do not necessarily indicate continuous occupation and do not track production during years between TPEs.

Population is considerably longer than in Dolores. There is some Basketmaker III and Pueblo I occupation contemporaneous with that in Dolores, followed by a large buildup of population during the prosperous Pueblo II period (ca. A.D. 950–1140). Of interest here, though, are the Pueblo III (ca. A.D. 1140–1280) occupations that terminate with the famous depopulation of the Four Corners region of the U.S. Southwest. The beginning of the Pueblo III period, in the mid-to-late 1100s, is poorly understood, with few tree-ring cutting dates, but during the 1200s there are several excavated hamlets and villages with well-dated timber procurement events (Fig. 4, bottom).

The general pattern is similar to that in Dolores: construction investment in hamlets is restricted almost completely to periods of average or high potential agricultural production. Construction investment in villages continues under all kinds of conditions, notably into the poor period that ends, ultimately, not only in the demise of the two villages contributing dated construction events to the graph but in the complete depopulation of the region. Here again we interpret the difference in pattern of cutting dates for low-investment hamlets versus high-investment villages as evidence of sunk-cost effects. The late Pueblo III period is also marked by considerable evidence for declining availability of protein in general and large game in particular and increased competition for the best agricultural land [Kohler 2000, Muir and Driver 2002, Varien, Van West, and Patterson 2000].

We draw a final, less conclusive example from an intermediate period. The most famous Puebloan struc-
tures, the “great houses” of Chaco Canyon, may show a similar pattern. Windes and Ford (1996) show that while early construction episodes in the early A.D. 900s in the canyon great houses Peñasco Blanco, Pueblo Bonito, Kin Bineola, and Una Vida typically coincide with periods of high potential agricultural productivity, later construction continues in both good periods and bad, particularly in the poor period ca. 1050–50, as might be expected in these high-investment settlements under the sunk-cost hypothesis. Unfortunately, there are very few cutting dates available for comparison from contemporary small sites in the canyon. Also complicating this example is the probability that relatively few people may have lived permanently in the canyon and that the great houses were used in part as centers for pilgrimage gathering in people from the much larger San Juan Basin (see, e.g., Renfrew 2001, Wills 2001). In effect this enhances the importance of i in the model and in turn would require that not only local but also regional productivity be considered. The productivity estimates used by Windes and Ford, generated by Rose, Robinson, and Dean (1996), are in fact based on a network of tree-ring series and are broadly applicable to the southeastern Colorado Plateau.

For one of these cases, that of the Sand Canyon Locality, there is a possible alternative explanation for continued investment in large sites during bad times. Widespread violence during the late 13th-century A.D. occupation of the northern Southwest (e.g., Kuckelman, Lightfoot, and Martin 2002) may have made life in small hamlets simply too dangerous. Some violence has been documented in 9th-century Dolores, but it seems to have been minor in comparison with the late-13th-century episodes, where the continued investment in villages may have been related as much to considerations of personal safety as to sunk-cost effects as we have defined them. However, the commonality of the patterns expected under the sunk-cost effect where no alternative explanation is evident, as in the Dolores case, suggests that this mechanism was effective in the histories of pre-Hispanic Puebloan societies.

In summary, in several cases we think we can discern evidence that people with large investments have, as a result of those investments, a tendency to attempt rather rigidly to maintain a previously successful way of life in areas and times when they are experiencing severely reduced returns on those investments—even to the point where they make additional investments in trying to maintain what perhaps ought to have been perceived as a lost cause. As a result, local decline becomes more severe than would have been the case had they chosen to leave earlier or otherwise changed the nature of their adaptation. In turn, the final collapse appears all the more dramatic, given the more impressive nature of the final structures left behind in a desolate landscape.

None of this should be understood to mean that we consider economic logic to be irrelevant in understanding prehistoric populations or prehistoric culture change in general. Rather, we suggest that sunk-cost effects be added to the catalog of factors suspected to make it difficult for populations to achieve optimal behavior in particular circumstances. Such a catalog would include uncertainty about the future, the high opportunity costs of obtaining more information, conflicting constraints, the low cost of following (possibly maladaptive) social norms, a tendency to follow (or, in some cases, the necessity of following) influential elites offering self-serving advice, and so forth. Somehow, despite these difficulties and distorting factors, economic logic, as a proxy for fitness logic, remains powerful for explanation of long-term trends. At smaller spatial and temporal scales, however, deviations such as those we think we see here help illuminate decision-making processes in particular societies.

Certainly, more archaeological records would need to be scrutinized to discover how general sunk-cost effects might be in inducing vulnerability to collapse in prehistoric societies with large structural investments. However, the strong empirical foundation in social psychological work and the fit to these Puebloan data encourage us to suspect that these effects are quite general. An attractive aspect of this model is that despite its simplicity it covers the three major ingredients that have been reappearing in the “collapse literature” for decades: the role of adverse events, the impressive size of collapsing settlements, and signs of overexploitation of resources during terminal occupations.

References Cited


JANSSSEN, MARCO A., AND MARTEN SCHEFFER. 2003. Overexploitation of renewable resources by ancient societies. MS.


Residential Mobility and Pottery Use in the Western Great Basin

JELMER W. EEKENS

Department of Anthropology, University of California, Davis, Calif. 95616-8223, U.S.A. (jeerkens@csulb.edu).

30 VI 03

The intersection of settlement patterns and material technology has long been of interest to archaeologists. For example, there is a huge body of literature on the relationship between mobility strategies and flaked stone assemblages. Similarly, much has been written on houses and other structures as they relate to mobility patterns. Unfortunately, for a range of reasons, similar effort has rarely been extended to ceramic technologies (but see Arnold 1985, Sassaman 1993, Bright and Ugan 1999, and Simms, Bright, and Ugan 1997).

At a general level, it is clear that pottery making and

The intersection of settlement patterns and material technology has long been of interest to archaeologists. For example, there is a huge body of literature on the relationship between mobility strategies and flaked stone assemblages. Similarly, much has been written on houses and other structures as they relate to mobility patterns. Unfortunately, for a range of reasons, similar effort has rarely been extended to ceramic technologies (but see Arnold 1985, Sassaman 1993, Bright and Ugan 1999, and Simms, Bright, and Ugan 1997).

At a general level, it is clear that pottery making and

The intersection of settlement patterns and material technology has long been of interest to archaeologists. For example, there is a huge body of literature on the relationship between mobility strategies and flaked stone assemblages. Similarly, much has been written on houses and other structures as they relate to mobility patterns. Unfortunately, for a range of reasons, similar effort has rarely been extended to ceramic technologies (but see Arnold 1985, Sassaman 1993, Bright and Ugan 1999, and Simms, Bright, and Ugan 1997).

At a general level, it is clear that pottery making and

The intersection of settlement patterns and material technology has long been of interest to archaeologists. For example, there is a huge body of literature on the relationship between mobility strategies and flaked stone assemblages. Similarly, much has been written on houses and other structures as they relate to mobility patterns. Unfortunately, for a range of reasons, similar effort has rarely been extended to ceramic technologies (but see Arnold 1985, Sassaman 1993, Bright and Ugan 1999, and Simms, Bright, and Ugan 1997).

At a general level, it is clear that pottery making and

The intersection of settlement patterns and material technology has long been of interest to archaeologists. For example, there is a huge body of literature on the relationship between mobility strategies and flaked stone assemblages. Similarly, much has been written on houses and other structures as they relate to mobility patterns. Unfortunately, for a range of reasons, similar effort has rarely been extended to ceramic technologies (but see Arnold 1985, Sassaman 1993, Bright and Ugan 1999, and Simms, Bright, and Ugan 1997).

At a general level, it is clear that pottery making and

The intersection of settlement patterns and material technology has long been of interest to archaeologists. For example, there is a huge body of literature on the relationship between mobility strategies and flaked stone assemblages. Similarly, much has been written on houses and other structures as they relate to mobility patterns. Unfortunately, for a range of reasons, similar effort has rarely been extended to ceramic technologies (but see Arnold 1985, Sassaman 1993, Bright and Ugan 1999, and Simms, Bright, and Ugan 1997).

At a general level, it is clear that pottery making and

The intersection of settlement patterns and material technology has long been of interest to archaeologists. For example, there is a huge body of literature on the relationship between mobility strategies and flaked stone assemblages. Similarly, much has been written on houses and other structures as they relate to mobility patterns. Unfortunately, for a range of reasons, similar effort has rarely been extended to ceramic technologies (but see Arnold 1985, Sassaman 1993, Bright and Ugan 1999, and Simms, Bright, and Ugan 1997).

At a general level, it is clear that pottery making and

The intersection of settlement patterns and material technology has long been of interest to archaeologists. For example, there is a huge body of literature on the relationship between mobility strategies and flaked stone assemblages. Similarly, much has been written on houses and other structures as they relate to mobility patterns. Unfortunately, for a range of reasons, similar effort has rarely been extended to ceramic technologies (but see Arnold 1985, Sassaman 1993, Bright and Ugan 1999, and Simms, Bright, and Ugan 1997).

At a general level, it is clear that pottery making and