Cultural responses to aridity in the Middle Holocene and increased social complexity

Nick Brooks*

Tyndall Centre for Climate Change Research, Saharan Studies Programme and School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

Abstract

The first complex, highly organised, state-level societies emerged in the Afro-Asiatic monsoon belt and northern South America during the 6th and early 5th millennia BP. This was a period of profound climatic and environmental change in these regions and globally, characterised by a weakening of the global monsoon system and widespread aridification in regions that today contain the bulk of the world’s warm deserts. This paper examines trajectories of socio-cultural and environmental change in six key regions in which complex societies emerged during the Middle Holocene: the central Sahara (focusing on the Libyan Fezzan), Egypt, Mesopotamia, South Asia (Indus–Sarasvati region), northern China and coastal Peru. Links between environmental and socio-cultural change are explored in the context of archaeological and palaeoenvironmental data and a theoretical framework of increasing social complexity as a response to enhanced aridity, driven largely by population agglomeration in environmental refugia characterised by the presence of surface water. There is direct evidence of adaptation to increased aridity in the archaeological literature relating to the Sahara and Egypt. In the other regions examined, the data are consistent with the notion that increased social complexity was largely driven by environmental deterioration, although further local-scale archaeological and palaeoenvironmental data are required to clarify the processes involved.

© 2006 Elsevier Ltd and INQUA. All rights reserved.

1. Introduction

Environmental catastrophes, particularly severe, rapid or abrupt changes in climate, are often associated in the academic literature with the collapse of civilisations. The fall of the Akkadian Empire and the end of the Egyptian Old Kingdom 4.2 thousand years before present (kyr BP) have both been attributed to climatic change resulting in a period of pronounced regional desiccation (Hassan, 1997; Weiss, 1997; Cullen et al., 2000). The collapse of the Tiwanaku state in the Titicaca Basin around 1 kyr BP has been interpreted as a result of drought (Ortloff and Kolata, 1993), as has that of the Mayan civilisation (Haug et al., 2003).

A widely held view is that the process of increasing ‘social complexity’ associated with the development of agriculture, large settled communities and the earliest states, was made possible by the relatively benign climate of the Holocene (e.g. Fagan, 2004; Burroughs, 2005). In this model of human social evolution, severe or abrupt changes in climate are associated with interruptions to or reversals in the progressive development of human societies. However, conversely, where uncalibrated dates, or years BC, are quoted directly from the original sources, they equivalent date in calendar years before present is given in parentheses, e.g. “…3300 yr BC [5.3 kyr BP].”

*Tel.: +441603 593900; fax: +441603 593901.
E-mail address: nick.brooks@uea.ac.uk.
there is widespread and increasingly abundant evidence that pronounced increases in social complexity in the Middle Holocene coincided with climatic and environmental deterioration, and in particular with increased aridity. This paper is concerned with the climatic and environmental changes, and the contemporaneous social and cultural developments, associated with the emergence of the earliest ‘civilisations’, defined here as complex, highly organised societies exhibiting a high degree of urbanisation, social stratification, specialisation of production, and centralisation of power. From here on the term ‘complex societies’ is used as a shorthand for such entities. It is not the purpose of this paper to address definitional issues relating to states and chiefdoms and the differences between them; here we are concerned with the emergence of the first societies characterised by large urban centres (i.e. the first ‘cities’) and exerting influence or control over large geographical areas, most probably through the exercise of some centralised power and associated formal institutions of governance. For detailed treatments of issues of complexity and ‘social evolution’ the reader is referred to Chapman (2003) and Yoffee (2005).

The first such complex societies emerged in the region currently occupied by the desert belt stretching from North Africa in the west to South Asia and China in the east, and in northern South America, during the late 6th and 5th millennia BP, a period characterised by a shift towards aridity in these regions. This paper examines the emergence of such societies in Egypt, Mesopotamia, South Asia (the Indus–Sarasvati region), northern China, and coastal Peru, and argues that their development was largely a response to increased aridity driven by global-scale changes in climate commencing around 6 kyr BP. However, the discussion of linked environmental and social change begins with the Sahara, where the evidence for large-magnitude changes in the physical environment and the associated impacts on human societies is relatively clear in comparison to other regions.

At this point it is worth stating some of the assumptions underpinning the following analysis. First, it should be stressed that the term ‘complex societies’ is used simply for convenience and brevity; the rigid classification of societies into successive evolutionary types representing ‘stages’ of social development is rejected here (see also Chapman, 2003; Yoffee, 2005). Secondly, the aim is not to argue in favour of environmental determinism or reductionism. Rather than seeing social and cultural change simply as being determined by changes or fluctuations in the physical environment, the view taken here is essentially a materialist, coevolutionist, one in which the physical environment is seen as setting the context within which social change occurs, providing both opportunities for, and constraints on, social, cultural, economic and technological innovation. Social and cultural contexts inform responses to environmental change, and are themselves influenced by the physical environment in which a society exists (Chapman, 2003).

One motivation for an examination of the role of the environment in the emergence of complex societies is the rather unsatisfactory nature of earlier models of increasing social complexity. These have invoked a wide range of causal factors in the development of what are popularly thought of as the first civilisations, including cultural diffusion from an original centre of innovation—usually Mesopotamia (e.g. Wheeler, 1968), increased agricultural productivity resulting in food surpluses capable of supporting non-producing members of society (e.g. Possehl, 2002), the need for central organisation of irrigation systems (Wittfogel, 1957), fortuitous technological innovation (e.g. Ratnagar, 2001), and the growth of trade networks (e.g. Possehl, 2002). Warfare has also been proposed as a driving factor in the emergence of state-level societies in other contexts (Spencer, 2003).

None of these factors provide a very satisfactory explanation for the emergence of the first complex societies when viewed in isolation, and it has been argued that many of the phenomena proposed as drivers of social complexity are more reasonably viewed as its products, albeit products that feed back into the process of increasing complexity (e.g. Fagan, 1999). More complex, subtle and multidimensional models, involving combinations of the above, have also been proposed (see Fagan, 1999; Chapman, 2003; Yoffee, 2005), but these do not satisfactorily explain the contemporaneous emergence of complex societies in widely separated parts of the world. The problem of explaining such closely spaced developments becomes more acute in the light of recent work that places the emergence of large urban centres in northern South America at the beginning of the 5th millennium BP (Solis et al., 2001; Mann, 2005).

Further impetus for a re-examination of the role of the physical environment in the emergence of the first complex societies comes from developments in our understanding of Holocene climatic and environmental change. It is increasingly apparent that the Middle Holocene was a time of profound change in environmental as well as cultural terms (Steig, 1999). In the early Holocene, the currently arid zone in the northern hemisphere extra-tropics was considerably wetter than today, and regions that are today covered by desert were well vegetated, supporting significant faunal and human populations. A wealth of palaeoenvironmental evidence, reviewed in more detail below, indicates that these regions became progressively more arid after the end of the 7th millennium BP, reaching a state similar to that existing today during the 5th millennium BP. Increasing social complexity, culminating in the emergence of highly organised state-level societies, parallels this process of environmental desiccation. Given the overwhelming importance of the physical environment in determining the suite of available livelihood options for prehistoric and proto-historic societies, it would be remarkable if such changes did not influence societal
development.

This paper argues that global climate change in the Middle Holocene stimulated societies to become more complex and organised as they responded to its local and regional manifestations. While this process is described in terms of broadly common responses to environmental change, the paper also addresses differentiated responses and diversity in outcomes resulting from the development of coupled social and environmental systems. While associations between social and environmental change in the Middle Holocene have been proposed in the past, to the author’s knowledge there has been no attempt to synthesise regional archaeological and palaeoenvironmental data with indicators of past climate change at the global scale. The increase in the availability of archaeological and palaeoclimatic data in recent years also puts such a synthesis on a firmer footing, although large gaps in our knowledge still remain.

The paper concludes with a brief discussion of the implications of the model of coupled socio-environmental change for our understanding of socio-cultural trajectories, and for current concerns about climate change and associated human adaptation.

2. Global and regional climate change in the Middle Holocene

2.1. The early Holocene humid phase and the 8 kyr BP arid interruption

The northern hemisphere arid belt, ranging from the tropics to the mid-latitudes in Africa, Asia and the Americas, and comprising most of the world’s warm deserts, was considerably wetter than today during the Early and Middle Holocene, from around 10 to 6 kyr BP (for a review see Lioubimsteva et al., 1998). Regions such as the Sahara were characterised by numerous water bodies and supported abundant humid-climate flora and fauna and significant human populations, and rainfall and surface water were more abundant throughout much of the northern hemisphere extra-tropical zone. Pachur et al. (1995) describe increased humidity in the early-middle Holocene in northern China as a manifestation of “a global strengthening of monsoon climate increasing precipitation from the Eastern Sahara to Inner Mongolia.” Palaeoclimatic data and modelling studies indicate that this “climatic optimum” was the result of differences in the Earth’s orbital parameters compared with the present day: increased seasonality and boreal summer insolation were associated with warmer ocean surface temperatures in the tropics, greater atmospheric moisture availability and intensified monsoon systems (Gasse and Van Campo, 1994; Kutzbach and Liu, 1997; Guo et al., 2000).

This warm humid period was punctuated by a number of cold episodes associated with enhanced aridity in the tropics and sub-tropics, the most severe of which occurred around 8 kyr BP (Gasse and Van Campo, 1994; Guo et al., 2000). Barber et al. (1999) associate this event with a cooling due to the collapse of a remnant of the Laurentide ice sheet, and the consequent injection of cold, fresh water via the Hudson Bay into the Labrador Sea. This would have resulted in a freshening of the North Atlantic, a suppression of deep water formation and a reduction in North Atlantic sea surface temperatures (Street-Perrott and Perrott, 1990). As well as reducing atmospheric moisture content and possibly inhibiting the development of rain-bearing convection systems, these changes would have significantly altered both the oceanic and atmospheric circulation; it is likely that all these factors contributed to a reduction in rainfall over regions such as northern Africa, which experienced a period of centennial-scale aridity around this time (Gasse and Van Campo 1994; Alley et al., 1997). The signal of the ~8 kyr BP event is evident across the globe, in the form of a widespread cooling signal in oxygen isotope records from Europe and Greenland between about 8.4 and 8 kyr BP (von Grafenstein et al., 1998), and evidence of an abrupt climatic reorganization lasting 200 years or less between 8200 and 7800 kyr BP at sites in equatorial East Africa, Antarctica and Greenland, as circulation underwent a transition to full postglacial conditions (Stager and Mayewski, 1997).

The ~8 kyr BP cold, arid event was an interruption to an otherwise humid period, followed by a return to wetter conditions throughout most of the northern hemisphere extra-tropical regions (Gasse and Van Campo, 1994). Such a recovery was made possible by the high levels of boreal summer insolation that led to the intensification of monsoon systems; during this part of the Holocene insolation alone was sufficient to cause the northward displacement of the monsoonal rainfall zone in the summer months. However, after 8 kyr BP boreal summer insolation gradually declined and it appears that monsoon dynamics became more sensitive to other factors, namely transient climatic perturbations and terrestrial feedbacks involving vegetation. Fleitmann et al. (2003) present the results of a study of an oxygen isotope record from a stalagmite in southern Oman, which reflects variations in the amount of monsoon precipitation from 10.3 to 2.7 kyr BP. From 10.3 to 8 kyr BP, decadal to centennial variations in monsoon precipitation are in phase with temperature fluctuations evident in Greenland ice cores. After 8 kyr BP, monsoon precipitation in southern Oman decreases gradually, indicating “a continuous southward migration of the mean summer ITCZ and a gradual weakening of monsoon intensity in response to declining June–August insolation at 30°N.” This is in contrast to more discontinuous changes apparent in other regions, where regional feedback processes led to abrupt changes. An example is the apparently rapid desiccation of the Sahara associated with a proposed collapse of the regional vegetation system which had acted to sustain the monsoon through moisture cycling in a weaker insolation regime (Claussen et al., 2003). It appears that after about 8 kyr BP, a robust, stable monsoonal regime driven primarily by solar insolation was replaced by a metastable regime prone to rapid and irreversible collapse.
This decoupling of the evolution of monsoon systems from solar insolation is important as it means that changes in monsoon rainfall are not likely to have been synchronous throughout the global monsoon belt. While the overall pattern of climatic change in the Middle Holocene was one of retreating monsoonal rainfall and associated environmental desiccation, there would have been variations in the local manifestations of this large-scale climatic change, and the nature and precise timing of regional or local episodes of desiccation would have been mediated by surface conditions that influenced local and regional climates within the wider monsoonal context. Environmental desiccation would have lagged climatic desiccation in some areas where water persisted at or near the surface as a result of local geological conditions. Where arid episodes were associated with external climatic shocks such as North Atlantic cooling events, their severity may also have depended on proximity to the Atlantic. Certain vegetation systems may have been more prone to collapse than others, while monsoon rainfall may have been more sensitive to global change further from oceanic moisture sources, such as in the Eastern Sahara. These factors mean that human populations throughout the monsoon belt would have been subjected to somewhat different desiccation trajectories as the rains retreated.

### 2.2. Enhanced aridity from 6 kyr BP

A wealth of evidence indicates that after 6 kyr BP, rainfall declined in the northern hemisphere sub-tropics and some of the adjacent extra-tropical regions, and that by ~4 kyr BP conditions similar to those pertaining today were established throughout much of this zone, with extreme aridity characterising most of the areas occupied by today’s extra-tropical deserts (Lioubimsteva et al., 1998). This transition appears to have been associated with a general global cooling and a retreat southwards of the northern hemisphere monsoon systems. In a review of Middle Holocene climate change, Steig (1999) describes 6 kyr BP as the time of steepest decline in northern hemisphere solar insolation since the early Holocene. It was during the period following 6 kyr BP that the first complex, urban, state-level societies emerged, in regions experiencing a shift towards aridity. The nature of the humid-arid transition in the sixth millennium BP is therefore of particular interest.

Just as the ~8 kyr BP cold, arid episode appears to have been associated with a transient cooling of the North Atlantic and an interruption of the early Holocene ocean dynamics, so a smaller-magnitude arid episode at around 6 kyr BP coincides with evidence of transient changes in the North Atlantic. Bond et al. (1997) conclude from an analysis of icerafted debris that cooling events occurred in the North Atlantic at 11.1, 10.3, 9.4, 8.1, 5.9, 4.2, 2.8 and 1.4 kyr BP, and a number of studies have concluded that these cold events were synchronous with cool, arid episodes and monsoon weakening at northern hemisphere low-mid latitudes during the Holocene (Kreutz et al., 1997; de Menocal et al., 2000; Berger and von Rad, 2002; Fleitmann et al., 2003; Gupta et al., 2003). For example, the 4.2 kyr BP event coincides with severe aridity in North Africa and Western Asia associated with the collapse of the Egyptian Old Kingdom and the Akkadian Empire (Hassan, 1997; Weiss, 1997; Cullen et al., 2000), while the 1.4 kyr BP event is associated with the Little Ice Age (de Menocal et al., 2000).

A dry episode centred around 6 kyr BP and approximately coinciding with the 5.9 kyr BP Atlantic cold event, followed by a partial recovery in some regions, appears to herald the onset of a period of discontinuous desiccation culminating in the creation of today’s global desert belt, extending from the Sahara to the Gobi and across southern North America. Guo et al. (2000) compare 560 radiocarbon dates from indicators of surface water in the Sahara with 158 dates from palaeosols and lake sediments from arid northern China and identify an arid episode occurring between 7 and 5.7 kyr BP, with a decrease in the frequency of humidity indicators after 5 kyr BP and the onset of extreme aridity after 4 kyr BP, in both regions. They conclude that “a dry interval did occur around 6 kyr BP”, and associate this with lowered sea-surface temperatures (SSTs) in the China Sea reported by Wang et al. (1995) and the end of a humid episode over the Arabian Sea described by Sirocko et al. (1993). Gasse and Van Campo (1994) review the evidence for arid and humid conditions at a variety of locations in Africa north of the Equator and in West Asia and find a dry episode recorded around 6 kyr BP at many, although not all sites. They find a recessional event at around 6 kyr BP at Lake Sumxi, western Tibet, a “stepwise return towards aridity” from around 6 kyr BP at Lake Bangong, also in western Tibet, “a brief recessional event at 6.0–5.8 kyr BP” at lakes Abhe’ and Ziway-Shala in Ethiopia, and dry events recorded in lake sediments at Termit and Bougdouma in the Sahel around 6.3 and 6.3–6 kyr BP respectively. They conclude that “several records ... suggest dry events of minor amplitude around 6 kyr BP,” in addition to “major dry spells ... recorded at all sites during the intervals 11.0–9.5, 8–7 and 3–4 kyr BP.” While “The 8–7 kyr BP dry spell interrupts a generally humid episode,” later dry periods “fingerprint the return towards aridity” (Gasse and Van Campo, 1994).

Reviewing lake records in Africa north of the Equator, Damnati (2000) states that “The majority of northern African lakes were at a high level at 6000 yr BP, although some had already begun to lower. The lakes show increasingly drier conditions after 6000 yr BP with the minimum lake levels being registered just after 4000 yr BP.” Specifically, Damnati (2000) identifies dry episodes in Mali, northern Chad and Sudan around and just before 6 kyr BP, and concludes that the decline in lake levels after 6 kyr BP was particularly pronounced between 10° and 22°N. This is consistent with a weakening and retreat of the African Monsoon, the strengthening of which had led to a northwards shift of the Sahara–Sahel boundary of some 500km relative to today.
prior to 6 kyr BP, to around 22–23°N (Lioubimsteva et al., 1998). Vernet and Faure (2000) present archaeological data that indicate a local minimum in evidence for human occupation in the Sahara south of 23°N at and just before 6 kyr BP, consistent with a failure of the monsoon at this time. Damnati (2000) sums up by stating that ‘‘6000 yr BP constitutes a transitional period, but was generally wetter than today.’’

These approximately synchronous changes in Africa and Asia are mirrored in palaeoclimatic records from extratropical North America. Baker et al. (2001a) present evidence for major environmental changes in the Upper Midwest of the USA at around 6 kyr BP, ‘‘when prairie rapidly replaced forest in the central Midwestern USA.’’ They suggest that changes in atmospheric circulation, perhaps coupled with changes in insolation, led to a shift in the boundary between different air masses and the replacement of ‘‘an essentially monsoonal circulation with a zonal circulation pattern.’’

Evidence for a weakened monsoon in the Americas is also presented by Haug et al. (2001), who use titanium and iron concentrations in anoxic marine sediments off the Venezuelan coast at ~10°N as proxies for variations in the hydrological cycle over northern South America, and find a trend towards drier conditions from ~5.5 kyr BP. Reflecting the conclusions of Fleitmann et al. (2003), they explain this trend as a result of shifts in the mean latitude of the Atlantic ITCZ.

2.3. Summary

The environmental response to reduced solar insolation after about 8 kyr BP was neither linear nor completely synchronous across the global monsoon belt. Arid episodes associated with Atlantic cooling events were superimposed on a gradual weakening of the northern hemisphere monsoon system. To reduced insolation and transient climate perturbations associated with Atlantic cold episodes we must also add terrestrial feedbacks associated with land-surface and vegetation-atmosphere interaction. These factors combined to cause a general trend towards enhanced aridity throughout the northern hemisphere extra-tropics, punctuated by more abrupt changes.

3. The central Sahara

Evidence for wetter conditions in the early and middle Holocene is particularly rich in the Sahara (Maley, 1977; Lezine, 1989; Lioubimsteva, 1995; Kutzbach and Liu, 1997; Jolly et al., 1998). Dating of archaeological sites, lake sediments and faunal remains indicates that wet conditions were established in the Sahara by around 10 kyr BP after a long period of aridity associated with the last glacial period (Goudie, 1992; Ritchie, 1994; Roberts, 1998). This humid phase, which lasted until around 5 kyr BP, with significant regional variations, was associated with an intensification of the African Monsoon caused by increased northern hemisphere summer insolation, resulting in its penetration far north of its current northernmost position (Ganopolski et al., 1998; Claussen et al., 1999, 2003; Tuenter et al., 2003).

The early Holocene humid phase was, however, punctuated by episodes of aridity that appear to have coincided with the North Atlantic cooling events evident in ice-rafted debris and Greenland ice-core records, mentioned above (Goodfriend, 1991; Alley et al., 1997 Bond et al., 1997; Smith, 1998; Guo et al., 2000; Cremaschi et al., 2001; di Lernia and Palombini, 2002). While summer insolation remained strong, the Saharan region recovered from these arid interruptions, and humid conditions were re-established. However, there is evidence that recovery was at best partial in the eastern Sahara after the arid event around 6000 BP (Goodfriend, 1991; Smith, 1998; di Lernia and Palombini, 2002). The entire Sahara had entered a period of desiccation by around 5 kyr BP (Cremaschi, 1998; Grandi et al., 1998; Jolly et al., 1998; Lioubimsteva, 1995). The process of environmental desiccation that followed the southward retreat of the monsoon was mediated by geography; while water persisted at or near the surface in some locations long after the cessation of significant rainfall, hyper-arid surface conditions were established rapidly in other Saharan regions (Cremaschi and di Lernia, 1998).

The climatic and environmental changes that affected the Sahara over the Late Pleistocene and Holocene were of such a magnitude that their impacts on human populations are in many respects very clear. Away from the Nile Valley and the coastal regions, the Sahara was essentially uninhabited during the last glacial period, when conditions were even more arid than at present, with the boundary between desert and savannah displaced some 500 km to the south relative to today, as evidenced by fossil dunes (Cremaschi and di Lernia, 1998; Talbot, 1983). The Sahara was reoccupied around or soon after 10 kyr BP; in central regions this reoccupation was by hunter gatherers following the monsoon rains as they extended northwards into the desert interior (Wasylikowa, 1993; di Lernia and Manzi, 1998; di Lernia, 2002). To hunting and gathering was added cattle herding, which spread westwards into the central and western Sahara after about 7 kyr BP.

The spread of cattle pastoralism appears to have been encouraged by climatic deterioration (Holl, 1998; Hassan, 2002; di Lernia and Palombini, 2002) describe a severe arid episode in the Acacus mountains of Libya in the late 7th millennium BP and an associated shift from semi-permanent settlements to seasonal migration. This episode may have been coincident with the cold, arid episode occurring around or before 6 kyr BP, and have resulted from the same causal factors. However, Nicoll (2004) presents abundant evidence for a drying trend in the deserts of Egypt and northern Sudan from the middle of the 7th millennium BP onwards, and it is possible that aridity affecting these areas extended to the Fezzan. It has also been suggested that the 8 kyr arid episode represented a shift from a year-round rainfall regime to a
seasonal monsoonal situation (Flohn and Nicholson, 1980), to which cattle herding (and the resulting predictability of food supply) would have been better suited than hunting and gathering. In order to understand the relationship between the environmental changes associated with the 8 kyr BP cold, arid episode and the spread of cattle herding, more numerous and precise radiocarbon dates are required; at present the resolution of the environmental and archaeological records is such that the full implications of decadal to centennial arid crises for Saharan populations cannot be elucidated in any great detail for this period (di Lernia, this volume).

Nonetheless, in the millennia following the 8 kyr arid interruption, cattle certainly became enormously important. In a more arid, variable or seasonal environment, cattle would have provided mobile pastoral groups with a more predictable source of nutrition in the form of milk, meat and blood than could be acquired from exploiting wild animal and plant species. Successful cattle pastoralism depended on the availability of pasture; recently it has been argued that the spatial distribution and dates associated with the evidence for cattle herding in the Sahara indicates a rapid spread westwards via discrete episodes of migration during arid crises in the 7th millennium BP, as herders were forced to move in search of new pasture (Hassan, 2002; di Lernia, this volume). This period saw the emergence of a pan-Saharan “cattle cult”, apparent from the ubiquitous depictions of cattle in rock paintings and engravings throughout the Sahara and the evidence for the ritualistic slaughter and burial of cattle (Sivili, 2002; Holl, 2004; di Lernia, this volume).

Of particular interest in the context of this paper is the Libyan Fezzan, where decades of research have illuminated trajectories of linked environmental and cultural change, which broadly reflect those of the Sahara as a whole (Cremaschi and di Lernia, 1998, 2001; di Lernia, 1999; di Lernia and Manzi, 2002; Mattingly et al., 2003a). The Fezzan is also instructive as it provides us with a model of increasing social complexity, sedentism and urbanization within a changing environment, in the form of the emergence of the Garamantian Tribal Confederation (Mattingly et al., 2003a, b). The following synopsis is based on di Lernia et al. (2002).

During the humid Early Holocene, livelihoods in the Fezzan, as throughout the Sahara, were focused on hunting and gathering. During the possibly more seasonal regime after the 8 kyr arid episode, hunting and gathering began to give way to cattle herding, although these two livelihood models coexisted for some time. As desiccation progressed, pastoralism based on sheep and goats, which are more tolerant or aridity, became more established; in the Acacus mountains and surrounding lowlands, the distribution of faunal remains suggests that in the late 7th millennium BP lowland pastures were set aside for winter cattle grazing, with sheep and goats being tended in highland regions. Desiccation was well established by around 5 kyr BP, and after this time cattle herding—previously the dominant livelihood activity—disappeared almost completely outside of the relict oases. di Lernia et al. (2002) describe the period centred on 5 kyr BP as the “hinge” between the Middle and Late Pastoral cultures of the Wadi Tanezzuft in the Libyan Fezzan, where they find a transition from animal to human burials. This reflects the situation in the Sahara as a whole, with the first monumental funerary structures, dated between 6.4 and 5.9 kyr BP, containing only faunal remains; this is followed by a mixture of human, faunal and empty/symbolic burials from 5.8 to 4.9 kyr BP, and finally by exclusively human burials from 4.8 kyr BP onwards (Sivili, 2002). In the Fezzan, di Lernia et al. (2002, p. 4) associate the increasing emphasis on human burials with “emerging figures within the pastoral group…” [representing] the first evidence … of increasing social stratification.” They suggest that funerary monuments also served to mark territory, boundaries, or zones of influence, formalising relationships between clan groups and the landscape as the region was subject to inward migration and increases in population density. Demographic changes resulting from environmental desiccation thus appear to have transformed social structures and led to innovation in the ideologies associated with death and burial. The distribution of monumental funerary architecture throughout northern Africa suggests that similar processes were unfolding throughout the Sahara.

In terms of livelihood strategies, human populations in the Fezzan responded to the post 5 kyr BP desiccation with a combination of increased sedentism and greater mobility. In higher elevation regions adjacent to the Wadi Tanezuft, the keeping of cattle was replaced by highly mobile pastoralism based on sheep and goats and involving large-scale year round movement in order to exploit remnant water and pasture, a nomadic lifestyle that persists to this day. In contrast, lower elevation regions were characterized by increasing settlement in relict oases, associated with sedentism and more intensive exploitation of local resources (di Lernia et al., 2002). The archaeological and palaeoenvironmental evidence suggests that this pattern of response was reproduced throughout the Fezzan, for example in and around the Wadi al-Hayat (al-Ajal), Wadi Barjuj and Wadi al-Shati (Mattingly et al., 2003a). In the regions dominated by these large wadis, environmental desiccation was not complete until the late 4th or early 3rd millennium BP. Cremaschi and di Lernia (2001) state that soil moisture reserves in the Wadi Tanezzuft were not fully depleted until about 3.5 kyr BP, while fluvial activity persisted until around 2.7 kyr BP. In the Wadi al-Hayat, in reality a scarp foot oasis rather than an ephemeral water course, there is evidence for the final drying of springs and the disappearance of surface water around 3 kyr BP, although some surface water may have remained in depressions now occupied by the playas at the lowest points of the Wadi al-Hayat (Brooks et al., 2003; Drake et al., 2004).

These dates for the final cessation of fluvial activity in the Wadi Tanezzuft and the termination of springs in the Wadi al-Hayat coincide with the beginnings of the Garamantian...
civilisation. The earliest evidence of settlement in the Wadi al-Hayat has been identified on the promontory of Zinkekra, accessed from the higher ground of the Messak Settafet and overlooking the Wadi al-Hayat, and has been dated to around 3 kyr BP (Mattingly, 2003a; Mattingly et al., 2003b). The Wadi al-Hayat was the primary location of later Garamantian settlement, as apparent from the many tens of thousands of burials along its southern edge, and was also the location of the principle Garamantian town of Garama (Old Germa), located at the edge of the present-day Germa Playa. Representing a local topographic minimum, and fed by groundwater rather than runoff, the Germa Playa would have been the last remaining significant water body in this part of the Sahara, and would have been a natural focus for human activity as the final stages of the regional desiccation unfolded. The situation of Garama is testament to this; it appears that the remaining human population congregated at the edge of the diminishing lake in a manner reminiscent of the situation on the shores of the West Nubian Palaeolake during the earlier desiccation of the eastern Sahara (Hoelzmann et al., 2001). Shallow wells have been excavated in the earlier layers of Old Germa, indicating that the earliest permanent settlement here occurred at a time when the local groundwater level was just below the surface (Drake et al., 2004). We may be confident that the rise of the Garamantes was a local development; di Lernia et al. (2002) state that “In a certain sense, Late Pastoral People became the Garamantes”, while Mattingly et al. (2003b) find some of the latest Pastoral lithics and pottery in the early Garamantian forts along the fringe of the Wadi al-Hayat.

At its height, Garamantian civilisation in the Wadi al-Hayat was based on irrigated agriculture made possible by the construction of foggara: inclined subterranean channels that tapped the elevated groundwater table at the base of the escarpment bounding the Wadi al-Hayat to the south, feeding water to the central areas of the Wadi (Wilson and Mattingly, 2003; Drake et al., 2004). This technology enabled the Garamantes to prosper and dominate the Fezzan from about 2.7 to 1.5 kyr BP, and challenge Roman hegemony in the Sahara (Mattingly, 2003b; Mattingly et al., 2003b). Rock paintings associated with the Garamantes are distributed widely throughout the central Sahara, extending into modern Algeria (Mattingly et al., 2003a). While we know little of their social organisation, it is clear that the Garamantes engaged in long-distance trade; the Garamantian heartland occupied a key location on trade routes between sub-Saharan Africa and the Mediterranean coast (Mattingly et al., 2003b). The construction and maintenance of the foggara in the loose alluvial and colluvial deposits of the Wadi al-Hayat would have involved considerable organisation. Numerous settlements and fortifications, stone architecture, quarrying activity, evidence of agriculture, abundant communal cemeteries, and of course the dense foggara networks, all point to a complex, organised society. Nonetheless, by the time of the first Arab incursions into the Fezzan, Garamantian society was in decline and was characterised by fragmentation (Mattingly et al., 2003b). The reasons for this decline are not known, but may include a fall in groundwater levels due to either the continuing lagged response to climatic desiccation or extensive exploitation, conflict with neighbouring populations, or economic stresses due to a decline in trade as a result of the collapse of the Roman Empire (Brooks et al., 2003; Mattingly et al., 2003b).

The Garamantian civilisation clearly appears to be the end result of a process of adaptation to increasingly scarce water resources unfolding over more than two millennia, and ultimately a response to the final disappearance of surface water around 3 kyr BP. The desiccation of the Fezzan led to population agglomeration, increased territoriality, social stratification, an increasing focus of human activity on specific locations, sedentism and technological innovation, leading to the emergence of a complex urban society. Although it emerged much later than the other complex societies dealt with here, there is a good case to be made for the Garamantian culture as a “pristine” civilisation, which arose from local cultural trajectories. While other factors may have assisted in the development of this civilisation—for example its location on emerging trade routes—environmental desiccation driven ultimately by climate change appears to have been the overwhelming force driving technological, economic, social and ideological change. The Fezzan of the Garamantes differs from other “cradles of civilisation” in that urbanisation did not occur along rivers; however, we might expect similar processes to be associated with population agglomeration in any refuge where water remains available during a time of desiccation.

4. Egypt

The cultural processes operating in the central Sahara as the environment became more arid are reflected in the archaeological record of the Nile Valley and the adjacent desert regions. In the very different, riverine environment of the Nile Valley, these processes culminated in the emergence of the Egyptian state and the unification of Egypt around or soon after 5.2 kyr BP (Grimal, 1992; Midant-Reynes, 1992). The fact that the Dynastic civilisation of Egypt emerged at a time of increasing aridity has been noted by a number of authors, who have linked increasing social complexity in the Nile region with changes in the physical environment (Midant-Reynes, 1992; Adams and Cialowicz, 1997; Malville et al., 1998). There is considerable evidence for environmental desiccation in the desert regions adjacent to the Nile Valley. Nicoll (2004) provides an extensive review of desiccation trajectories the eastern Sahara, finding that most rainwater-fed playas in Egypt wane after about 7 kyr BP, and are significantly desiccated by 5.5 kyr BP; the most northerly playas begin to dry as early as 7.7 kyr BP. Throughout the 7th millennium BP, numerous hydrological indicators point to progressive desiccation. By around 6 kyr BP full desert conditions are evident in southern Egypt except in some oases and wadis, and “Hand-dug wells at
many of the playa sites after 6000 BP suggest that the onset of arid conditions and commensurate falling water tables influenced human activities” (Nicoll, 2004, p. 571).

Wendorf and Schild (1998), working at Nafta Playa in south west Egypt write that “Around 6200 yr cal BP the modern phase of hyperaridity began in the Eastern Sahara and the area [around Nafta Playa] was abandoned.” In the higher elevation region of the Gilf Kebir, to the north west of Nafta Playa, Linstädter and Kröpelin (2004) find evidence for an “unparalleled climatic transition at about 5500 [14C] yr BP/4400 yr BC [6.4 kyr BP]” when a summer monsoonal rainfall regime was replaced by one of winter rainfall, heralding the terminal phase of the Holocene pluvial. During this phase the plateau of the Gilf Kebir was used by “groups with nomadic pastoral economies, until even these last retreat areas had to be abandoned due to the final desiccation of the Eastern Sahara around 4500 [14C] yr BP (3300 yr BC) [5.3 kyr BP]” (Linstädter and Kröpelin, 2004, p. 775).

Hoelzmann et al. (2001) find that in the northwest of present-day Sudan, the level of the West Nubia Palaeolake was declining in the 6th millennium BP, with evidence of settlements increasingly apparent on lower ground at the fringes of the lake as the shoreline retreated. They find evidence for intensive use of the region around the lake shore by cattle rearing groups between 5.9 and 4.4 kyr BP (5200 and 4000 14C yr BP), after which the area was abandoned.

The overall pattern of human–environment interaction in the far east of the Sahara appears to be one of pastoral groups following increasingly sparse and localised rainfall from the end of the 7th millennium BP, congregating in highland areas and around shrinking water bodies. It is likely that at least some of these groups would have headed towards the Nile Valley as these last refuges dried during the final phase of desiccation in the Eastern Sahara in the late 6th millennium BP. Malville et al. (1998, p. 448) suggest that “an exodus from the Nubian Desert at 4800 [14C] years BP [5.6 kyr BP] may have stimulated social differentiation and cultural complexity in pre-dynastic Upper Egypt.” Such a process could also have been stimulated by a limiting of the subsistence options for groups that already spent at least some the year in the Nile Valley itself: Wilkinson (2003) argues that populations that had previously practiced seasonal migration between the Nile Valley and the summer savannah in what is now Egypt’s Eastern Desert were forced to settle permanently in the Nile Valley as a result of the cessation of summer rainfall.

Midant-Reynes (1992, p. 232) places the cultural trajectories of late Predynastic times within an ecological context, describing a “gradual movement of human settlements from the deserts towards the river valley” throughout the 6th millennium BP, which accelerated during the final stages of the Predynastic period “…bringing with it the relative abandonment of pastoralism, and the adoption of intensified agriculture, backed up by increasingly systematic artificial irrigation.”

Linked processes of environmental and cultural change are particularly evident at Hierakonpolis (where Nekhen, the capital of prehistoric Upper Egypt was situated), whose population was gradually squeezed into the confines of the alluvial plain, abandoning the desertified wadis. Here, as at Elkab [on the opposite, east bank of the Nile], the alluvial deposits of the ‘Nekhen formation’ came to an end in around 3200BC [5.2 kyr BP], at around the same time as the last local phase of the Holocene pluvial. (Midant-Reynes, 1992, p. 232, citing Hoffman et al., 1986).

Wengrow (2001, p. 97) finds evidence at Hierakonpolis suggesting “an attempt to combine the maintenance of herds with increased sedentism, probably through an increase in artificial feeding with cultivated grain,” further supporting the hypothesis that social change was associated with the environmentally driven settlement of pastoral groups in the Nile Valley due to declining water and pasture in the neighbouring deserts.

The emergence of Egyptian Dynastic civilisation, and the unification of Egypt, thus occurs at a time of significant environmental, and associated social, change. Increased population density resulting from a decrease in habitable or productive land in the late Predynastic period is associated with increased social stratification and the appearance of monumental architecture in the form of temples and large, sometimes elaborately decorated, tombs (Hoffman, 1971, 1982; Midant-Reynes, 1992; Rice, 2003), mirroring developments in the Sahara as discussed above. An emerging elite controls the trade in raw materials and a class of skilled workers appears which achieves elevated social status through its association with the “royal” authority of the early Pharaohs (Midant-Reynes, 1992). The increase in population density drives innovation in agricultural technology as modes of food production are adapted to a new socio-ecological environment, and also provides an impetus for the eventual expansion of the Naqada culture of Upper Egypt that leads to the unification of the country around 5.2 kyr BP (Midant-Reynes, 1992). Adams and Cialowicz (1997, p. 57) explicitly link the Naqada expansion and the consequent formation of the Pharaonic state to environmental change, stating that the expansion of the Upper Egyptian territory was “encouraged by the pressure of a greater population in the south, where climatic change in the late Predynastic had reduced winter rainfall and husbandry in the deserts and brought about a reliance on agriculture in natural basins.”

The emergence of the Egyptian Dynastic state from the earlier prehistoric cultures may be viewed as a result of adaptation to increased aridity throughout the 6th millennium BP, followed by a period of rapid desiccation and associated demographic change in the late PreDynastic period, when competing “proto-state entities” coalesced in Upper Egypt (Maisels, 1999). Certain groups would have been at a relative advantage over their contemporaries as a result of
their geographical location and economic situation, providing a context within which social stratification and the emergence of ‘royal’ power could proceed (Midant-Reynes, 1992). Migrant groups arriving in the Nile Valley are likely either to have come into conflict with existing populations or to have formed disadvantaged groups, either of which would have increased social stratification further. Groups of lower status would have provided a pool of labour which could have been exploited by the emerging elites for military purposes and the monumental building projects that were to become a prominent feature of Egyptian society. While the nature of the Naqada expansion and the unification of Egypt is still a matter of some debate—i.e. was it peaceful or warlike, characterized by conquest or assimilation? (Midant-Reynes, 1992) - the longer term processes that led to the emergence of the unified Egyptian state can be understood best when viewed in their wider environmental context.

5. Mesopotamia

Ancient Mesopotamia is generally viewed as the location in which the first complex, urban, state level societies emerged during the Uruk period in the 6th millennium BP, the result of a long transition from village communities to large integrated estates and finally cities (Hole, 1994; Matthews, 2003). The first evidence for social stratification appears around 7 kyr BP at the beginning of the Ubaid period, when the size of some settlements increases and monumental architecture in the form of temples, centralized storage facilities and elite residences appear (Hole, 1994). According to Hole (1994, p. 122)

These innovations take place in the context of a demographic crash which began about 5500 BC [7.5 kyr BP]. Settlement in some regions disappears, and in others declines rapidly. No region shows an increase in sites and no region was newly opened to settlement. The pattern of gradual expansion across the arable landscape of the near east by small villages [0.5–2 ha in size], begun by 7700BC [5.7 kyr BP], had ended. Now an increasing proportion of people resided in the larger towns and much of the landscape was devoid of settlement.

However, the extent and nature of social stratification and differentiation is a matter of some debate (Stein, 1998; Matthews, 2003). The Ubaid was characterised by a significant level of cultural uniformity, evident from pottery and architectural styles, indicating significant regional interaction. The presence of exotic materials such as lapis lazuli and carnelian, and the widespread use of stamp seals, indicate long-distance trade and the development of administrative systems to control the exchange of sealed goods (Hole, 1994; Matthews, 2003).

The transition to the Uruk culture after about 6 kyr BP was associated with a dramatic increase in the number of settlements, some of which grew into towns and cities by the middle of the 6th millennium BP; by the late Uruk period towards the end of the 6th millennium BP some of the larger settlements were walled (Pollock, 1999). Matthews (2003, p. 108) states that “...most researchers would probably agree that in the Uruk period, through more of less the entire fourth millennium BC, the first true states appeared. These states, moreover, were primary states, originating in pristine condition on the plains of Mesopotamia.” The Uruk period is associated with the mass production of pottery, the style of which differed significantly from the Ubaid, the further development of accounting systems introduced during or before the Ubaid, the introduction of writing, and explicit representations of violence and authority (Matthews, 2003). The Uruk culture is generally viewed in terms of an expansion from a core region around the city of Uruk-Warka in southern Mesopotamia, with evidence of Uruk material culture being found in the outer regions of Mesopotamia, often in Uruk “enclaves”. There is considerable debate as to the nature of the Uruk expansion; while some scholars view this phenomenon in an imperial context, others question the view of Uruk as an imperial capital (Algaze, 2001; Matthews, 2003).

The Uruk period came to an end around 5.2 kyr BP, and was followed by the Jemdet Nasr period, which lasted for no more than two centuries and was characterised by regional differentiation and the abandonment of many Uruk settlements outside southern Mesopotamia (Pollock, 1999; Matthews, 2003). The principal Uruk city of Uruk-Warka nonetheless grew dramatically, and “The region around Uruk-Warka played host to a sudden tenfold increase in settlement density at about 3200 BC [5.2 kyr BP], coupled with the development of a four-tiered hierarchy of settlement, all made possible by increased availability of dry and very fertile land newly freed from constant inundation by an ameliorating [sic] climate” (Matthews, 2003, p. 110, citing Nissen, 1988, pp. 66–67). This period also appears to have been associated with increased conflict and the fortification of settlements, suggesting competition for resources as much as abundance resulting from the availability of new fertile land (Leick, 2001, p. 55; Schwartz, 2001, p. 262). It was followed by a period characterised by the interactions between competing and cooperating city states, and ultimately the emergence of the Akkadian and Babylonian empires (Pollock, 1999; Leick, 2001).

Trajectories of environmental change in Mesopotamia are more complex and less well understood than those of the central Sahara and Egypt. In particular, the courses of the Tigris and Euphrates were more variable than that of the Nile, and shifts in river courses had dramatic impacts on local environments, meaning that caution must be exercised in palaeo-environmental interpretation (Maisels, 1999; Matthews, 2003). Changes in sea level in the Gulf region of southern Mesopotamia must also be considered (Hole, 1994). Nonetheless, attempts have been made to establish chronologies of environmental change, although such reconstructions are based largely on evidence from regions
adjacent to the southern Mesopotamian alluvial lowlands, where palaeo-environmental data are extremely scarce. Hole (1994) explicitly places the development of urbanisation in an environmental context. He argues that a humid period from 12 to 6 kyr BP provided favourable conditions for agriculture, while changes in sea-level caused the frequent relocation of settlements in southern Mesopotamia, preventing the development of large integrated socio-agricultural systems and urban centres. He associates the onset of the Ubaid period with a “rapid decline from favourable climatic conditions in the second half of the sixth millennium BC [8th millennium BP] [that] resulted in substantial dislocations of populations and abandonment of some regions,” based on evidence from pollen records in southern Mesopotamia and evidence from the Zagros mountains presented by El-Moslimany (1983) and Kay and Johnson (1981). As during the later desiccation in the Sahara and Egypt, people responded through a combination of increased mobility and sedentism in favourable locations.

Hole (1994, p. 123) argues that the Uruk expansion is associated with a more stable physical environment that “encouraged the development of more extensive canal systems and larger communities”. This view is compatible with that of Nissen (1988), who suggests that drier conditions in southern Mesopotamia in the 6th millennium BC made more land available for agriculture. These arguments reflect the view that the development of complex societies is facilitated of a benign environment that enables human society to progress, referred to at the beginning of this paper. However, Hole (1994) also cites Sirocko et al. (1993), who presents evidence from Arabian Sea sediment cores indicating a rapid increase in the amount of airborne mineral dust around 5.5 kyr BP. Similar evidence is used by Cullen et al. (2000) and de Menocal (2001) to infer a link between severe aridity and the later collapse of the Akkadian empire at 4.2 kyr BP. Such a link is also postulated by Bar-Matthews et al. (1999) based on an analysis of oxygen and carbon isotope ratios extracted from a speleothem in Soreq Cave near the Mediterranean coast of Israel, in which the record of δ18O values between 4.1 and 4 kyr BP are indicative of aridity. They find even higher δ18O values between 5.2 and 5.1 kyr BP, suggesting more severe aridity at the end of the Uruk period, at the time of the large increase in population density around Uruk-Warka (Matthews, 2003). Cullen et al. (2000) also find an increase in the amount of aeolian dust transported to the Gulf of Oman at 5.2 kyr BP; however, their results indicate that this is not as pronounced as that at the end of the 5th millennium BP. These findings are supported by evidence from the regions adjacent to southern Mesopotamia, which suggests a precipitation minimum from around 5.2–4.9 kyr BP (Butzer, 1995, cited in Pollock, 1999). Wright (2001) also reports that the wider region was subject to significant environmental change during the 6th and 5th millennia BP, characterised by increasing aridity.

The cultural trajectories evident in Mesopotamia are very different to those in the Sahara and Egypt. Moves towards urbanisation and “complexity” started earlier in Mesopotamia, with the emergence of larger villages and the use of accounting systems during the Ubaid period. In Mesopotamia, a degree of cultural uniformity over large geographic areas during the Ubaid and Uruk periods gives way to increased regional differentiation at the beginning of the Jemdet Nasr period, commencing around 5.2 kyr BP. In Egypt the opposite occurs, with the unification Upper and Lower Egypt around this time, after a period of regionalism characterised by interacting proto-state entities. However, these developments appear to happen within similar climate change contexts, involving instabilities in rainfall patterns following the global cold (and in these regions, arid) episode around 8 kyr BP. By the 6th millennium BP a process of aridification had set in, and in the middle to late 6th millennium BP this process appears to have accelerated as severe decadal to century-scale droughts had pronounced impacts on the physical environment and on human societies. In particular, both Egypt and Mesopotamia seem to have experienced abrupt environmental and social changes around 5.2 kyr BP, presaging similar changes at 4.2 kyr BP which coincided with the collapse or transformation of social systems as attested by written records (Trigger et al., 1983; Rice, 2003). Evidence for common environmental and cultural experiences in Egypt and Mesopotamia during the cold episode at 4.2 kyr BP further suggests that the late 6th millennium BP is likely to have been a time of profound environmental change in Mesopotamia as well as in the Nile Valley and surrounding deserts (Cullen et al., 2000; de Menocal, 2001).

This evidence suggests that the social changes in the Late Uruk period and at the transition to the Jemdet Nasr period were the result of crisis—specifically episodes of severe drought spanning as much as a century or more—rather than internally generated social dynamics associated with the exploitation of a stable environment. As in the central Sahara and the Nile Valley, it seems that people evolved new ways interacting with one another and with the physical environment as they gathered around environmental refugia, resulting in often dramatic and rapid increases in population density. Algaze (2001) has postulated a combination of changes in river courses and increasing aridity as drivers of increasing social complexity, stimulating social instability, regional competition and conflict, and population agglomeration. These processes appear to have been operating in Mesopotamia since the 8th millennium BP, and to have accelerated towards the end of the 6th millennium BP, encouraging what Hole (1994) refers to as the “urban revolution.”

6. The Indus–Sarasvati region

The arid region straddling the borders of present-day India and Pakistan saw the flowering of what is commonly referred to as the Indus or Harappan civilisation in the 5th
millennium BP. This culture flourished along the Indus and Ghaggar-Hakra (or ancient Sarasvati) river valleys, and might more accurately be referred to as the Indus–Sarasvati civilisation.

There is evidence of climatic and environmental desiccation during the 6th millennium BP in this part of South Asia, reflecting the situation in northern Africa and western Asia. In the Thar desert, to the east of the Indus Valley, sediments from Lake Lunkaransar reveal fluctuations in the water table in response to the early Holocene Indian Monsoon with the same decadal to centennial scale variations as apparent in records from the Arabian Sea (Burns et al., 1998; Enzel et al., 1999). Enzel et al. (1999) conclude from the sedimentary record that Lake Lunkaransar, whose level is controlled by groundwater and local runoff, did not dry between the beginning of the Holocene and the middle of the 6th millennium BP (10 and 4.8 14C ka), and experienced a high stand between about 7 and 5.7 kyr BP (6.3 and 4.8 14C ka). A major and unprecedented (within the context of the Holocene) environmental change led to an abrupt fall in lake levels around 6.4 kyr BP (5.5 14C ka), and the lake was completely dry by around 5.5 kyr BP (4.8 14C ka). The decline in water levels coincides with an increase in sand transport to the lake, indicating a destabilisation of the surfaces of the surrounding dunes.

The situation in the Thar desert as recorded in the Lake Lunkaransar record is reflected to the east in the Ganga plain, where Srivastava et al. (2003) describe the conversion of river channels into ponds between 8 and 6 kyr BP, with fluvial activity in the region ceasing sometime between 7 and 5 kyr BP. Srivastava et al. (2003) describe the hydrological conditions of the Ganga plain as being controlled largely by climatic variability associated with the monsoon rains, supported by tectonic activity. Schuldenrein et al. (2004) examine the alluvial histories of sites in the vicinity of the ancient city of Harappa and find evidence for the stabilization of landscapes in the early Holocene during optimal climatic conditions of stable rainfall and moderate evapo-transpiration rates. This was followed by some destabilization in the Middle Holocene, apparent as evidence of renewed river channel migration and thinner Middle Holocene soils. Schuldenrein et al. (2004, p. 795) conclude that: “A gradual turn to desiccation, ca 7–6 kyr BP, may be signalled by resumption of loess sedimentation in the Ghaggar-Hakra plain, reduction in the scale of alluviation and incision in some of the main river valleys, and stabilization (and initial lowering) of lake levels.”

Enzel et al. (1999, p. 127) point out that the regional desiccation described above preceded the emergence of the urban phase of Indus–Sarasvati civilisation by many centuries, and that: “The Indus civilization flourished mainly along rivers during times when northwestern India experienced semiarid climatic conditions that are similar to those at present.” As was the case for the earliest complex civilisations of Egypt and Mesopotamia, the Indus culture arose once the regional climate and environment had become more arid.

The Indus–Sarasvati civilisation, like the contemporary cultures of Mesopotamia and Egypt, had its roots in much earlier cultures. The earliest evidence of agriculture in the greater Indus region is attested at Mehrgarh, in the far west, around 9 kyr BP, lying within what McIntosh (2002) refers to as the Western Asiatic “interaction sphere”. Possehl (2002, p. 29) presents a chronology of what he terms the “Indus Age”, in which a period represented by the “beginnings of village farming communities and pastoral camps” from around 9 to 6.3 kyr BP (Stage 1) is followed by one of “developed village farming communities and pastoral societies” from around 6.3 to 5.2 kyr BP (Stage 2). This is succeeded by the “Early Harappan” (Stage 3) from 5.2 to 4.6 kyr BP, followed by a rapid transition to urbanisation (Stage 4) over a period of about a century which ushered in the “Mature Harappan” phase associated with large urban centres such as Harappa, Mohenjo-Daro, Chanhu-Daro, Kalibangan and Lothal.

Possehl (2002, p. 34) describes Stage 2 as “a special time in the prelude to the Mature Harappan”, associated with technological innovation, changes in burial practices, and some evidence of population movement. The later centuries of Stage 2 see an expansion of farmers and herders eastwards from Baluchistan into the Punjab, where they occupied the plains of the Indus and Sarasvati rivers, particularly “the eastern drainage of the ancient Sarasvati River” (Possehl, 2002, p. 35). Herders also appeared at the eastern fringes of the Thar Desert, and Stage 2 also brings a process of regionalisation, inferred from distinct regional archaeological assemblages (Possehl, 2002).

Viewed within an environmental context, regionalization and the movement of populations to the river valley regions are reminiscent of the processes of linked environmental and social change evident in Egypt and proposed above for Mesopotamia. Occurring after the onset of the process of desiccation that affected the entire arid belt of Africa and Asia, it is reasonable to suggest that these social changes may have been influenced by environmental change, particularly given the evidence for regional aridification, summarised above. As in Egypt and Mesopotamia, the inferred weakening of the monsoon system coincided with a focusing of human activity around rivers, where water resources would have been more reliable and environmental variability more predictable, resulting in greater food security.

The Early Harappan (Stage 3) was characterised by a continuation of village farming communities and pastoralism, with the latter engaging in transhumance between highland regions in summer and lowland floodplains in winter, penetrating deeper into the riverine zones in search of pasture during this stage (Possehl, 2002). Such a pattern of landscape exploitation by mobile pastoralists reflects the situation in the Sahara prior to its final desiccation around 5 kyr BP. McIntosh (2002) places
the beginnings of urbanization around 5.2 kyr BP, at the beginning of the Early Harappan, and the earliest archaeological deposits at Harappa date from around 5.3 kyr BP (Belcher and Belcher, 2000).

Nonetheless, Possehl (2002, p. 46) concludes that there is “little evidence for significant social differentiation, craft and career specialization, and little evolution of the political and ideological institutions that produce public architecture” in Stage 3. These traits appeared in the Mature Harappan (Stage 4), along with large-scale urbanisation, writing and a system of weights and measures, town planning, bureaucracy, social stratification, and the apparent emergence of the state and political differentiation (Possehl, 1990). The transition from Stage 3 to Stage 4 was extremely rapid, occurring within the space of about a century in the middle of the 5th millennium BP. McIntosh (2002) places this transition between around 4.7 and 4.6 kyr BP, with Possehl (2002) suggesting a later transition between 4.6 and 4.5 kyr BP. There is agreement about the highly discontinuous nature of the transition, during which “the haphazard settlements of the Early Indus people disappeared” to be replaced by planned towns and cities built on their ruins (McIntosh, 2002).

There was widespread abandonment of sites during the Stage 3–Stage 4 transition, and many sites exhibit evidence of large fires, suggestive of a common pattern of destruction, the causes of which are the subject of controversy (McIntosh, 2002; Possehl, 2002). Many sites were abandoned in this transitional period; although new, planned settlements were built on the ruins of some earlier settlements, many Stage 3 settlements were abandoned completely, and a large number of Stage 4 settlements were built on previously unoccupied land. Possehl (2002, p. 50) presents data from Sindh and Cholistan; in Cholistan 33 of 37 recorded Stage 3 sites were abandoned completely, while 132 of the 136 recorded Stage 4 sites were on “virgin soil”. In Sindh, a similar, although less dramatic, pattern is apparent.

Possehl (1990) contrasts this “paroxysmal” change with the more continuous trajectories towards urbanisation in Mesopotamia, with its “clear lines of continuity and change for the core features of Sumerian civilization”. While acknowledging Indus–Sarasvati cultural continuity, he states that “the Harappan Civilization created a distinctive set of signs and symbols that can easily be differentiated both from what came before it and from the material culture of the contemporary peoples in adjacent regions” and suggests that the revolutionary change of the mid-5th millennium BP may have been driven by a set of ideological precepts. He also suggests that trading contacts with Mesopotamia (most likely interacting with ideological shifts and other factors) may have driven or influenced these changes, by leading to a re-evaluation of resources, enhancing the potential for social stratification and trade specialisation, and increasing tendencies towards population agglomeration which led to the development of new forms of social control.

While ideological change may be inferred from changes in material culture and iconography, and while there is evidence of trade between the Indus–Sarasvati and Mesopotamian civilisations, there is no direct evidence that these factors drove the dramatic changes associated with rapid urbanisation in the mid-5th millennium BP. Neither is there any direct evidence that this transition was driven by environmental change. Nonetheless, there are echoes of the environmentally embedded cultural trajectories of Egypt and Mesopotamia in the archaeology of the greater Indus–Sarasvati region. Evidence of population movements and a focusing of activity around river valleys after the onset of long-term desiccation throughout the desert belt of the Old World around 6 kyr suggests that the populations of the greater Indus–Sarasvati region may have been responding to environmental change. The beginning of the Early Harappan (Stage 3) coincides approximately with the emergence of the unified Egyptian state and the transitional Jemdet Nasr period in Mesopotamia, although in the Indus–Sarasvati region social change is not so dramatic at this time. Rapid, discontinuous change in this region occurs some half a millennium later with the abrupt transition to urbanisation.

The reasons for the dramatic social changes that led to the emergence of the Mature Indus–Sarasvati civilisation are unknown. Could trading contacts with Mesopotamia, or ideological or religious innovations have precipitated such dramatic changes, or do we need to look to some more forceful external stimulus? Such a stimulus could conceivable have taken the form of rapidly increasing aridity that drove pastoral groups to settle permanently along the river valleys, swelling the populations in these areas and precipitating widespread social change, as appears to have happened in Egypt (albeit in a less discontinuous fashion). Such an event may have nurtured or accelerated ideological innovation; trade and associated social differentiation and stratification may also have played a role in the transition to urbanisation. Whether organised trade, formal religious ideology, and social stratification and differentiation are the driving factors behind the emergence of complex civilisation, or more likely products of it, is a matter of some debate (Fagan, 1999). Once the transition to such a society is underway, we can be confident that these factors will interact, and it is the nature of this interaction, mediated by the broader environmental context, that gives a complex society its unique identity.

7. Northern China

The emergence of the first state-level societies in China has been placed at the end of the 5th to the beginning of the 4th millennium BP (Liu, 1996; Lee, 2004). These emerged from the Longshan cultures on the central plains of northern China, described by Liu (1996, p. 237) as “the crucial matrix in which the first states evolved from the basis of earlier Neolithic societies.” The Longshan period, commencing around 4.8 kyr BP and following the Yangshao cultural
period in Henan and Shanxi provinces, was associated with "a process of social change from more egalitarian to stratified societies" (Liu, 1996, p. 242). These cultures were not comparable in terms of complexity to those discussed previously; Lee (2004, p. 187) concludes that "the pattern of settlement during the Longshan phase is consistent with that of chiefdoms, particularly that of simple chiefdoms." Lee's (2004) analysis is based on data from the Lower Yi-Luo River Valley Survey, which indicates that during the Late Longshan period, and indeed prior to the end of the 3rd millennium BP, settlements in this part of northern China were strongly clustered along river courses. Lee (2004) interprets the presence of two larger settlements (13 and 20 ha in size compared with smaller settlements averaging 1.7 ha) as an indication of emerging social complexity. Again it must be noted that the Longshan culture was not comparable in terms of urbanisation, organisation or stratification with the other societies discussed here; in terms of settlement size distribution it is more reminiscent of Mesopotamia during the Ubaid period, some two millennia earlier.

As was the case in the other regions discussed above, the middle Holocene in northern China was a time of profound climatic and environmental change. A wealth of evidence indicates a general transition from a warm, humid climate to cooler, arid conditions across the region, with the exact nature and timing of this transition varying with location (An et al., 2000; Zhang et al., 2000; Yang et al., 2003). Wei and Gasse (1999) conclude from an analysis of lacustrine carbonates from five lakes in the Qinghai-Tibet Plateau and North Xinjiang in northwest China that an abrupt intensification of summer rainfall occurred at 12–11 kyr BP, "associated with an extremely rapid 1000–1500 km westward migration of the monsoon fronts. The monsoon regime weakened in stages after 8–7 cal. kyr BP [and] Holocene maximal aridity is recorded at all sites between about 4.5 and 3.5 cal. kyr BP."" Zhang et al. (2000) describe a climatic sequence for the Hongshui River in the southern Tengger Desert of northern China, north west of the present monsoon limit, consisting of an unstable climate in the early Holocene followed by a more stable, warm climate from the middle of the 8th until the end of the 6th millennium BP. They identify a shift towards aridity in the early 6th millennium BP, followed by an abrupt drop in temperature some centuries around 5.3 kyr BP which heralds the beginning of a stepwise deterioration in climate over the following millennia, culminating in an extreme dry spell and the final cessation of fluvial-lacustrine activity around 3 kyr BP. Xiao et al. (2004) analyse past vegetation changes as recorded in sediments from Lake Daihai in north-central China at the present summer monsoon limit and find evidence for cooling and enhanced aridity around 5.1 kyr BP, followed by an amelioration around 4.8 kyr BP and a further cold arid period between 4.5 and 4.0 kyr BP, after which climatic fluctuations were superimposed on a general trend towards a cooler, drier climate until around 2.9 kyr BP.

Liu (1996) associates the Yangshao–Longshan transition with climatic deterioration at the 6th–5th millennium BP transition, and also proposes a correlation between other cultural transitions in northern China and changes in both climate and river courses. In Central Henan province, Liu (1996) infers (from the moderate increase in settlement numbers over time) that the population was relatively stable during the Yangshao period. After the Yangshao period the population multiplied threefold over the following 800 years; in a process reminiscent of that proposed by some authors for the increase in population in southern Mesopotamia in 6th millennium BP. Liu (1996) suggests that this was a result of migration to the stabilizing landscape of the Yellow River Valley from adjacent regions as the climate became drier. As in the other regions discussed above, increased aridity would have resulted in the river valleys becoming the main focus of activity.

8. Peru

Recent archaeological work in coastal Peru has demonstrated the existence of a complex state-level society in the fifth millennium BP, described as "one of the world's biggest early urban complexes" (Mann, 2005, p. 34). The ruins of this civilisation, located in a collection of river valleys in an area known as the Norte Chico region, consist of more than 20 separate residential centres and exhibit monumental architecture, in the form of pyramid-like structures dominating the urban landscapes (Haas et al., 2004). This area is situated around 11°S, and so lies well outside the latitudinal range of the Afro-Asiatic zone with which we have been concerned so far. Nonetheless, Mann (2005, p. 34), reviewing the findings of archaeologists working in the Norte Chico region, writes that "The concentration of cities in the Norte Chico is so early and so extensive, the archaeologists believe, that coastal Peru must be added to the short list of humankind's cradles of civilisation, which includes Mesopotamia, Egypt, China and India." Given the theme of this paper it is therefore appropriate to address the emergence of this South American culture alongside the more well-known early civilisations of Africa and Asia.

The broad cultural trajectory of the Peruvian coast reflects those of the other key regions discussed above. Stanish (2001) writes that "At the beginning of the fourth millennium B.C.E. [Before the Common or Christian Era: i.e. the 6th millennium BP] all peoples in South America lived in small hunting, gathering, and horticultural camps, or, on rare occasions, in small semipermanent villages. By 3000–2500 B.C.E., [5–2.5 kyr BP] the first fully sedentary and complex societies developed on the Pacific coast of Peru." Solis et al. (2001, p. 723) deduce from radiocarbon dates from the site of Caral in the Supe Valley of the Norte Chico region that "monumental corporate architecture, urban settlement, and irrigation agriculture began in the Americas by 4090 [14C] years before the present (2627 calibrated years BC) to 3640 [14C] years before the present (1977 calibrated
years BC) [4.6 to 3.8 kyr BP].’’ The Caral data are revised to ‘‘about 2800 B.C.E.’’ or 4.8 kyr BP in the later review article by Mann (2005). Solis et al. (2001) describe extensive structures likely to have had ceremonial and/or administrative functions in the central zone of Caral, which covers some 65 ha. Formally arranged residential complexes are also apparent, and these features are reflected at other sites in the Supe Valley.

These findings place the emergence of complex societies in the Americas within a few centuries of their counterparts in the Afro-Asiatic desert belt. The large number of unexcavated settlement mounds in the Peruvian coastal region mean that earlier dates cannot be excluded. Given the lack of contact with the emerging civilisations of Egypt and Asia (unless we heed some of the more outlandish theories that exist at the fringe of archaeology), these independent developments in the Americas are remarkable, and further suggest that increasing complexity in the Middle Holocene was the result of independent responses to some external stimulus.

Evidence for Middle Holocene climatic and environmental change in tropical South America is abundant; however, the signals are complex and often contradictory. For example, Cross et al. (2000) present evidence that water levels in Lake Titicaca were up to 100m lower than at present between 6.8 and 4.1 kyr BP (6 and 3.8 ¹⁴C kyr BP), a conclusion bolstered by Abbott et al. (2003). Rigsby et al. (2001) report that major downcutting episodes in a southwestern tributary of Lake Titicaca most probably occurred between about 6 and 4.5 kyr BP, during periods of decreased precipitation on the Altiplano and decreasing levels of Lake Titicaca. Baker et al. (2001b) conclude that maximum aridity and lowest lake levels occurred between 8 and 5.5 kyr BP. However, Placzek et al. (2001) describe this period as one of high water levels in Lake Aricota, only 130 to the southwest of Lake Titicaca, and sharing a similar climatology. Similarly, Holmgren et al. (2001) do not find any evidence for a Middle Holocene drought near Arequipa, 200km west of Lake Titicaca in the northern Atacama Desert. The spatial extrapolation of palaeoclimatic data in the tropical Andean region must therefore be approached with extreme caution.

Nonetheless, there is reasonable evidence for changes in the coastal environment in north-central Peru in the Middle Holocene. Reitz and Sandweiss (2001) detail changes in marine vertebrate resources at Ostra Base Camp, a Middle Preceramic site occupied between 6250 and 5450 BP and situated some 200km north of the Supe Valley, which they argue indicate a change from a tropical to temperate marine ecology around the latter date. They review data from other sites that support their conclusion that ‘‘The shift from tropical to temperate taxa seems most likely to be due to a change in the water conditions prevailing along the coast prior to 5000 BP north of 10ºS latitude.’’ They suggest that this change was associated with the onset of ‘‘an ENSO cycle of long intervals of cold water conditions and brief warm water incursions,’’ in the late 6th millennium BP, replacing a regime of more-or-less permanently warm water associated with either very short intervals between El Niño and La Niña phases, or much longer-duration warm phases associated with El Niño conditions. Work in the same location by Andrus et al. (2004) indicates that summer sea surface temperatures (SSTs) in this region were around 3ºC warmer than those of the present day, with mean annual temperatures being some 3–4ºC warmer, during the middle and late 7th millennium BP. Andrus et al. (2004) conclude from a variety of data that coastal upwelling associated with the Peru–Chile current intensified after about 5 kyr BP.

Coastal upwelling, in addition to its impacts on marine ecosystems, also exerts an influence on terrestrial climate; cold water delivered by upwelling off the western coasts of South America and southern Africa causes coastal aridity through its cooling and stabilising impact on the overlying atmosphere (Hastenrath, 1991). It is therefore plausible that Middle Holocene changes in ocean temperature/circulation, evident in the marine records of coastal Peru, resulted in enhanced aridity in coastal areas, as well as in significant changes in the marine resource base. We can propose a plausible hypothesis in which these changes reduced the viability of food procurement strategies based solely on marine resources, forcing communities to diversify their livelihood strategies, combining marine and terrestrial resources. Wells and Noller (1999) suggest that populations along the Peruvian coast became progressively more dependent on terrestrial resources in the Middle Holocene, and that ‘‘The shift to an agricultural economy resulted in a migration of settlements inland along the river valleys. Extreme events (sea level stabilization, droughts, El Niño floods) have likely facilitate periods of rapid technological and cultural innovation.’’ Evidence from the Supe Valley suggests that coastal and inland settlements in the 5th millennium BP were interdependent, with sites such as Caral generating agricultural produce, for example cotton, that was traded with coastal settlements such as El Aspero (Mann, 2005). Such a complex symbiotic relationship, essentially one of geographically differentiated specialisation, might have been encouraged by a need or desire to exploit both coastal and inland resources resulting from a shift to a less productive coastal environment. Such a model is highly speculative at this stage, however, and must remain as a hypothesis to be tested by future research. We must also bear in mind the significant distance between the key sites referred to in this discussion: Ostra, providing evidence of environmental change, and the Supe Valley, suggesting complex interactions between coastal and inland regions; we cannot automatically assume that conditions at one of these locations pertained in the other.

9. Discussion

It is now well established that the earliest complex, highly organised, state-level societies emerged at a time of increasing aridity throughout the global monsoon belt. This
trend towards desiccation commenced around 8 kyr BP as a result of declining solar insolation associated with changes in the Earth’s orbital parameters, but accelerated around 6 kyr BP after a widespread centennial-scale arid episode that may have been the result of transient cooling in the North Atlantic. The following millennium was a time of profound cultural change that saw the development of the world’s first states in Mesopotamia and Egypt, and laid the foundations for similar developments in South Asia, northern China and northern South America. The available data suggest further abrupt changes in climate in the late 6th millennium BP, when regional records indicate environmental and cultural discontinuities. The period around 5.2 kyr BP seems particularly significant. At this time a unified Egyptian state emerged, and the Uruk culture of Mesopotamia collapsed and gave way to the transitional Jemdet Nasr period, characterised by fragmentation and regionalism. In South Asia the beginnings of the Early Harappan phase have been placed at 5.2 kyr BP (Possehl, 2002) and the beginnings of urbanisation are evident at the site of Harappa (McIntosh, 2002). In northern China there is evidence for an abrupt drop in temperature and accelerated aridity coinciding with the Yangshao–Longshan transition in the final centuries of the late 6th millennium BP (Liu, 1996). In South America this period was characterised by profound changes in the ENSO cycle and an increase in coastal upwelling (Reitz and Sandweiss, 2001; Andrus et al., 2004) prior to the emergence of large urban centres exhibiting monumental architecture. Changes in the behaviour of monsoon systems appear to be one aspect of a wider reorganisation of the global climate.

The association between environmental desiccation and increasing social complexity is particularly striking in the central Sahara and Egypt. Trajectories of environmental change are quite clear in the Sahara, and are based on a variety of proxy data from a large number of locations. Geoarchaeological studies in the Libyan Fezzan tell a coherent story of demographic and cultural change driven predominantly by changes in water availability, although different adaptive responses are evident. In Egypt there is abundant evidence of the abandonment of the deserts flanking the Nile Valley, increased population densities in the Nile Valley itself, and attempts to transform lifestyles rooted in mobile cattle pastoralism to ones suited to a more sedentary existence in a confined geographical area. The model of competing “proto-state entities” eventually coalescing in the late 6th millennium BP is compatible with the expected impacts of the process of desiccation that began around 6 kyr BP and culminated around or soon after 5 kyr BP. This desiccation, and the resulting scarcity of water and pasture, eventually would have led to the collapse of livelihoods based on mobile pastoralism in the Western and Eastern deserts of Egypt. Those pastoral populations that did not perish would have migrated to refugia such as the Nile Valley, either with their animals in search of water and pasture, or as what we might term “environmental refugees” as a last resort. Increases in population density, competition over resources and trade routes, and the emergence of social strata associated with the presence of advantaged and marginalised groups (for example, established groups and migrants from the surrounding deserts) are likely to have provided the conditions for the development of political power and competition between political entities.

The links between environmental and social change are less well defined in the other regions discussed in this paper, and are to a large extent inferred by comparing distinct archaeological and regional palaeoenvironmental records. In the case of Mesopotamia, palaeoenvironmental data in the southern alluvial lowlands are scarce, and interpretation of past climatic and regional environmental change is further complicated by the role of changes in river channels in local environmental change. Nonetheless, there is abundant evidence of regional environmental desiccation and, while the trajectories of social change in Mesopotamia and Egypt are very different, in both cases increasing social complexity is associated with increases in local population densities, the congregation of populations along rivers, and evidence of competition or conflict at the end of the 6th millennium BP. In Egypt this occurs within a context of political unification, whereas in Mesopotamia the context is one of fragmentation. In both cases environmental deterioration coincides with trends towards greater urbanization and the development of ideologies of political power.

The transition to highly urbanised societies exhibiting the characteristics of states occurs later in the Indus–Sarasvati region and northern China. In both regions increasing social complexity is associated with a concentration of settlements along rivers at a time of increasing aridity, followed by the emergence of large urban centres. The transition to urbanisation was particularly rapid and discontinuous in the Indus–Sarasvati area, and strongly suggests a time of social upheaval during which many existing settlements were deliberately destroyed. Whatever the role of the environment in this transition, there were clearly other processes at work, although their precise nature (e.g. violent conflict and/or radical ideological change) is unknown.

Evidence for increased social complexity during periods of regional climatic and environmental change in the Middle Holocene is not restricted to the Afro-Asiatic desert belt. It appears that similar developments were occurring in coastal Peru in the early 5th millennium BP. Reconstructing environmental trajectories in this region is complicated by the high level of heterogeneity in the palaeo-environmental record, associated with geographical and topographic factors. Nonetheless, responses to increased coastal aridity provide a convincing explanation for the cultural developments evident in the archaeological record, and are consistent with what is known about the nature of Middle Holocene environmental change in this region.

Explanations of increasing social complexity involving environmental change will remain controversial until (and probably even if) they can be supported by further archaeological data. In particular, they must ultimately be
assessed through field studies that examine linked socio-cultural and environmental trajectories at the local scale. This will require numerous studies of individual settlement sites in order to illuminate the nature of changes in livelihood strategies, social organisation and ideological frameworks, supported by palaeoenvironmental proxy data from the vicinity of the settlements in question. While regional palaeoenvironmental data can provide us with a broad context within which to interpret cultural change, they tell us little or nothing about the nature of the interactions between people and the physical environment at the scales which would have been important to the communities from which complex societies arose.

While much more work needs to be done on the links between environmental deterioration and the emergence of complex, organised, state-level societies, the case for climate change induced aridification as the principal driving force behind the development of the first “civilisations” is strong. The evidence for dramatic and sometimes abrupt climate change in the Middle Holocene is overwhelming, and we would expect such changes to have profound impacts on societies that were directly dependent on their immediate physical environment for their livelihoods and food security. Changes in this physical environment would have resulted in changes in the available options for and constraints on livelihood strategies. We may view social change within such a context through the prism of adaptation to climatic and environmental change, mediated by other factors such as pre-existing socio-economic relations, livelihood strategies, and ideologies.

A recurring theme in this discussion has been population agglomeration in environmental refugia, necessitating the development of new social institutions and relations, and technological and institutional adaptations related to water extraction, food production and distribution. As suggested by Fagan (1999), these developments are not necessarily dependent on the prior emergence of an organising elite as proposed by Wittfogel (1957); it might be argued that the emergence of elite groups is a by-product of, rather than a prerequisite for, such adaptations, and results from the exploitation by certain groups of both geographic advantage and emerging social relations and institutions that provide opportunities for exerting control over other, relatively disadvantaged, social groups. People arriving in refugia would have been more likely to constitute such disadvantaged groups, providing a pool of “human capital” available for exploitation through organised labour or military activity by emerging elites. Ideological systems developed to legitimise the power of elites and support the emerging social hierarchy (Yoffee, 2005), while innovations in administration, production, and distribution would have been required in order to maintain the emerging social system. In the model presented here, increases in social complexity leading to the emergence of urbanization and state-level societies are not driven by surpluses in food production. Instead, the capacity of agricultural systems to generate surpluses (from which high-density, specialised, urban societies can be supported) is harnessed out of necessity. While surpluses may have played a role in the first steps towards complexity, for example in the Ubaid and earlier periods in Mesopotamia and in the development of village agriculture in pre-Harappan times in South Asia, the rapid increases in complexity in the late 6th and early 5th millennia BP are interpreted as precipitated by hardship rather than abundance. This interpretation reflects the main conclusion of Marshall and Hildebrand (2002) regarding the domestication of plants and animals in Africa, which they argue was driven by a desire for greater predictability in the food supply rather than a deliberate attempt to increase yields and thus volumes of available foodstuffs. They point out that “The end results of agriculture—visible today as larger yields, higher carrying capacity, denser stands of crops, larger seeds and seed heads, or greater animal productivity—were not necessarily realized during the earliest phases of the domestication process” (Marshall and Hildebrand, 2002, p. 101). As with agriculture, so with urbanisation and the development of the institutions of state—the consequences of such developments would not have been anticipated by those populations that unwittingly precipitated them.

It should be stressed that this paper addresses the development of a specific set of societies at a particular period in time; its purpose is to illuminate the processes associated with the emergence of these particular societies and to draw some very broad general lessons rather than to identify any universal laws of human-environment interaction. The emergence of complex, urban, state-level societies is not an inevitable outcome of environmental deterioration; in the Middle Holocene, increased mobility proved in many respects to be a more sustainable response to aridity than technological innovation and urbanization (di Lernia and Palombini, 2002). Neither should we assume that urbanisation and state formation only occur as a response to environmental change. There are numerous other instances of state formation, and it is not proposed that the model elaborated here applies to all of them. Conversely, there are instances of urbanisation during the Early Holocene in western Asia that did not culminate in the development of complex, state-level societies (Akkermans and Schwartz, 2003).

Furthermore, the diversity of the societies discussed above, and the different trajectories leading to them, should caution us against reductionism. While environmental desiccation provided the context for, and was arguably the principal driving force behind, their emergence, a multitude of other factors influenced their development. Each complex society discussed here emerged from a different cultural context. For example, the Garamantes ultimately emerged from the cattle herding cultures of the Sahara, and mobile cattle-based societies appear to have formed a component of proto-Dynastic Egyptian society; there was no significant tradition of agriculture in the prehistoric Sahara. This may be
contrasted with the long-established and geographically widespread tradition of village agriculture in Mesopotamia, and livelihoods based on marine resources on the pre-urban Peruvian coast. The different physical environmental contexts would have provided each precursor society with different constrains and opportunities as they became more complex. Ideological systems would have developed with reference to the physical environment, but would also have been influenced by personalities and by contacts with other cultures. Trade would also have played a role in shaping livelihoods, in the development of economic and social structures, and in shaping world views. Each of the mature complex societies discussed here was distinct from the others, exhibiting its own unique ‘personality’; the unified Egyptian state in which the bulk of the population remained rural may be contrasted with the much more urbanised society of Mesopotamia, characterised during much of its existence by competing and cooperating city states. The high levels of social stratification in Egyptian and Mesopotamian societies, dominated by powerful ruling elites and individual leaders, suggest very different social structures to the apparently far less hierarchical Indus culture (McIntosh, 2002).

In terms of responses to ‘abrupt’ climate change, it should be noted that the climatic changes apparently associated with the emergence of complex societies are qualitatively similar to those blamed for societal collapse. In the case of Egypt, increased aridity is associated with the emergence of the unified Dynastic state in the late 6th millennium BP and with the widespread collapse of social systems during the First Intermediate Period some thousand years later at 4.2 kyr BP. In Mesopotamia, aridity appears to be associated with the emergence of the Uruk culture, its subsequent collapse and the transition to the Dynastic period, and also with the later collapse of the Akkadian Empire. Transient environmental shocks thus appear to have a variety of outcomes depending on the nature of the societies on which they impact, and the relationship between environmental and socio-cultural change is not straightforward, with a given environmental phenomenon having a single type of societal outcome. For example, it has been argued that the development of agriculture, which formed the basis for the increases in social complexity in the Middle Holocene discussed here, was associated with a succession of responses to both climatic amelioration and deterioration (e.g. Hole, 1991).

The model of linked environmental change presented here has a number of implications. Firstly, it provides another challenge to linear models of social evolution in which societies evolve through a number of predefined developmental states. There is a significant element of convergent evolution in the societies examined here, and they emerged as a result of increasing complexity which might be interpreted as taking them through ‘stages’ of development. However, the trajectories that led to their emergence exhibit considerable diversity and discontinuity, associated to a large extent with their regional environmental contexts. In this model, increases in complexity are not a result of the inevitability of progress, but are driven by responses to changes in the physical environment on which people depend for their livelihoods. Social complexity may therefore be viewed largely as emerging from adaptation during the Middle Holocene, although it would clearly be nonsense to claim that complexity is only ever driven by adaptation to environmental change. Furthermore, not all adaptation in the Middle Holocene leads to increased social complexity as defined here; responses involving increased mobility can be at least as successful as those based on urbanisation. The social systems emerging from adaptation to aridity based on mobile pastoralism are in some respects more resilient to further changes than their complex and in many respects fragile urban counterparts, as apparent from the persistence of pastoralism long after urban civilisations have disappeared (di Lernia and Palombini, 2002).

The resilience of responses based on mobility holds important lessons for the emerging fields of adaptation research and policy, which are developing in response to concerns about current and future anthropogenic climate change. In today’s globalising world, traditional livelihoods are under pressure from economic liberalisation, monetisation of local economies and development programmes based largely on western models. Factors such as drought and conflict can also threaten traditional systems. Given the uncertain nature of future changes in climate in semi-arid regions such as the Sahel (Brooks, 2004), the building of flexible and resilient livelihoods is a priority. The creation of ‘enabling environments’ for adaptation and the building of ‘adaptive capacity’ have been stressed as priorities in order to facilitate adaptation (Brooks and Adger, 2005); support for traditional livelihoods that have evolved to cope with climatic variability is one way of achieving this. Any measures to enhance resilience and promote adaptation must take account of local contexts and recognise that imported developmental models may be inappropriate; indigenous livelihood strategies have often emerged from centuries or even millennia of linked environmental and social change.

The archaeological record can offer more general lessons about adaptation to climatic and environmental change. Today, adaptation is seen as a means of minimising or neutralising the adverse impacts of anthropogenic climate change, of preserving existing social systems, and of supporting predefined developmental goals. In this sense, adaptation is seen as something that is manageable and, by implication, predictable. This is evident in the language of adaptation research and policy, which talks of ‘adaptation goals’ and the design of ‘adaptation strategies, policies and measures’ (e.g. UNDP, 2005). While such approaches are necessary to confront the consequences of climatic changes that are the inevitable results of past and unavoidable near-future greenhouse gas emissions, they may be unrealistically optimistic. The archaeological record emphasises that
adaptation has in the past been associated with great social upheavals that could not have been foreseen by those who were undertaking the adaptation. The consequences of adaptation were unplanned and unpredictable, arising from the ad hoc responses of a variety of actors to environmental change. Planned adaptation strategies designed to help societies cope with anticipated future changes in climate represent a radical shift in the way human beings respond to and interact with the physical environment. Furthermore, the view of adaptation as a means of neutralising the impacts of environmental change is a naïve one, particularly when such change is abrupt in nature. Rather than being the result of deliberative processes, past adaptations have as a rule emerged from upheavals triggered by environmental change. Past adaptation has occurred out of necessity, after damages have already been incurred, and itself is not without cost. We may be justified in viewing civilisation as a form of adaptation to climate change, but the negative aspects of the transitions with which its development is associated, for example increased inequality and violent conflict, suggest that it should perhaps be viewed as a “suboptimal” adaptation by the standards of present-day aspirations.

10. Conclusions

The archaeological and palaeoenvironmental evidence is consistent with the notion that the development of complex societies in the Middle Holocene was largely the consequence of the responses of the precursor societies to deteriorating environmental conditions. This deterioration was associated principally with the orbitally driven weakening and southward retreat of the northern hemisphere monsoon belt. In all the regions examined in the preceding sections, the emergence of complex societies coincided with or followed a period of increased aridity. A general trend towards desiccation after about 8 kyr BP was punctuated by shorter (decadal to centennial scale) episodes of increased aridity. While desiccation trajectories were mediated by local and regional factors and feedback processes such as the collapse of vegetation systems, there appear to have been episodes of accelerated aridification that were coherent throughout the monsoon belt. One such event occurred around 6 kyr BP, and may have been associated with cooling in the North Atlantic. Enhanced regional aridity following this event coincided with sociocultural change, particularly in the Eastern Sahara, Egypt and Mesopotamia. Regional data suggest another episode of accelerated change at the end of the 6th millennium BP, when discontinuities are apparent in archaeological records from across the Afro-Asiatic desert belt, and abrupt environmental changes are suggested by records from northern Africa, Western Asia, China and northern South America.

The evidence for linked environmental and social change is very strong in the central Sahara and Egypt, where responses to aridity are evident in local archaeological records. While the data are consistent with the hypothesis that social complexity was stimulated by increased aridity in Mesopotamia, the Indus–Sarasvati region, northern China and coastal Peru, further field-based research is required in order to link social change explicitly with environmental change at local scales, for example as represented by individual settlements. In particular, more high resolution palaeoenvironmental data are required from these regions in order to establish local trajectories of environmental change that may be related to local archaeological records.

References


Hoffman, M.A., 1982. The Predynastic of Hierakonpolis: An Interim Report. Egyptian Studies Association, Publication No. 1, Cairo University Herbarium, Faculty of Science, Giza, Egypt, and the Department of Sociology and Anthropology, Weltern Illinois University, Macomb, IL.


