

from atmospheric chemists from around the world.

First, establish and fund a network of 'flagship' stations<sup>9</sup> to monitor: concentrations, fluxes, interactions and feedbacks as well as more general air quality and meteorology data. Around 5–8 such stations (costing between US\$7 million and \$11 million each) would suffice for a major city. These should be complemented by mobile measurement platforms on cars and aeroplanes, remote sensing of air columns from the ground, satellite observations and smog chambers. Major sources of pollutants can be identified using historic data.

Second, indoor air-quality measurements and monitoring must be conducted concurrently in a representative selection of residential and office buildings.

Third, atmospheric chemists must model secondary-pollutant production pathways and feedback mechanisms under high concentrations of various pollutants. These models must then be compared with observations.

Fourth, the links between air pollutants and mortality and other health effects need to be established. That way the most health-relevant pollutants and their sources can be identified and mitigated first. A database should be developed to track health impacts.

Fifth, long-term sustainable engineering solutions such as improving processes and material flows in industry must be implemented to maintain low levels of air pollution. This will require capacity building across the Chinese authorities and institutes on using air-quality assessment data in decision-making, in developing legislative tools and in clean-air action plans.

Only by understanding atmospheric chemistry will China clean its air. ■

**Markku Kulmala** is professor of aerosol physics at the University of Helsinki, Finland.

e-mail: markku.kulmala@helsinki.fi

1. Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D. & Pozzer, A. *Nature* **525**, 367–371 (2015).
2. Lim, S. S. *et al. Lancet* **380**, 2224–2260 (2012).
3. Zhang, J. *et al. Lancet* **375**, 1110–1019 (2010).
4. Guo, S. *et al. Proc. Natl Acad. Sci. USA* **111**, 17373–17378 (2014).
5. Ding, A. J. *et al. Atmos. Chem. Phys.* **13**, 5813–5830 (2013).
6. Ding, A. J. *et al. Atmos. Chem. Phys.* **13**, 10545–10554 (2013).
7. Xiao, S. *et al. Atmos. Chem. Phys.* **15**, 1769–1781 (2015).
8. Kulmala, M. *et al. Annu. Rev. Phys. Chem.* **65**, 21–37 (2014).
9. Hari, P. *et al. Atmos. Chem. Phys. Discuss.* **15**, 21063–21093 (2015).



The mysterious Indus unicorn on a roughly 4,000-year-old sealstone, found at the Mohenjo-daro site.

# Cracking the Indus script

**Andrew Robinson** reflects on the most tantalizing of all the undeciphered scripts — that used in the civilization of the Indus valley in the third millennium BC.

The Indus civilization flourished for half a millennium from about 2600 BC to 1900 BC. Then it mysteriously declined and vanished from view. It remained invisible for almost 4,000 years until its ruins were discovered by accident in the 1920s by British and Indian archaeologists. Following almost a century of excavation, it is today regarded as a civilization worthy of comparison with those of ancient Egypt and Mesopotamia, as the beginning of Indian civilization and possibly as the origin of Hinduism.

More than a thousand Indus settlements covered at least 800,000 square kilometres of what is now Pakistan and northwestern India. It was the most extensive urban

culture of its period, with a population of perhaps 1 million and a vigorous maritime export trade to the Gulf and cities such as Ur in Mesopotamia, where objects inscribed with Indus signs have been discovered. Astonishingly, the culture has left no archaeological evidence of armies or warfare.

Most Indus settlements were villages; some were towns, and at least five were substantial cities (see 'Where unicorns roamed'). The two largest, Mohenjo-daro — a World Heritage Site listed by the United Nations — located near the Indus river, and Harappa, by one of the tributaries, boasted street planning and house drainage worthy of the twentieth century AD. They hosted the world's first known toilets, along with ▶

► complex stone weights, elaborately drilled gemstone necklaces and exquisitely carved seal stones featuring one of the world's stubbornly undeciphered scripts.

### FOLLOW THE SCRIPT

The Indus script is made up of partially pictographic signs and human and animal motifs including a puzzling 'unicorn'. These are inscribed on miniature steatite (soapstone) seal stones, terracotta tablets and occasionally on metal. The designs are "little masterpieces of controlled realism, with a monumental strength in one sense out of all proportion to their size and in another entirely related to it", wrote the best-known excavator of the Indus civilization, Mortimer Wheeler, in 1968<sup>1</sup>.

Once seen, the seal stones are never forgotten. I became smitten in the late 1980s when tasked to research the Indus script by a leading documentary producer. He hoped to entice the world's code-crackers with a substantial public prize. In the end, neither competition nor documentary got off the ground. But for me, important seeds were sown.

More than 100 attempts at decipherment have been published by professional scholars and others since the 1920s. Now — as a result of increased collaboration between archaeologists, linguists and experts in the digital humanities — it looks possible that the Indus script may yield some of its secrets.

Since the discovery of the Rosetta Stone in Egypt in 1799, and the consequent decipherment of the Egyptian hieroglyphs beginning in the 1820s, epigraphers have learnt how to read an encouraging number of once-enigmatic ancient scripts. For example, the Brahmi script from India was 'cracked' in the 1830s; cuneiform scripts (characterized by wedge-shaped impressions in clay) from Mesopotamia in the second half of the nineteenth century; the Linear B script from Greece in the 1950s; and the Mayan glyphs from Central America in the late twentieth century.

Several important scripts still have scholars scratching their heads: for example, Linear A, Etruscan from Italy, Rongorongo from Easter Island, the signs on the Phaistos Disc from the Greek island of Crete and, of course, the Indus script.

In 1932, Flinders Petrie — the most celebrated Egyptologist of his day — proposed an Indus decipherment on the basis of the supposed similarity of its pictographic principles to those of Egyptian hieroglyphs. In 1983, Indus excavator Walter Fairservis at the American Museum of Natural History in New York City, claimed in *Scientific American*<sup>2</sup> that he could read the signs in a form of ancient Dravidian: the language

**"No firm information is available about its underlying language."**



Mohenjo-daro existed at the same time as the civilizations of ancient Egypt, Mesopotamia and Crete.

family from southern India that includes Tamil. In 1987, Assyriologist James Kinnier Wilson at the University of Cambridge, UK, published an 'Indo-Sumerian' decipherment, based on a comparison of the Indus signs with similar-looking ones in cuneiform accounting tablets from Mesopotamia.

### THREE PROBLEMS

In the 1990s and after, many Indian authors — including some academics — have claimed that the Indus script can be read in a form of early Sanskrit, the ancestral language of most north Indian languages including Hindi. In doing so, they support the controversial views of India's Hindu nationalist politicians that there has been a continuous, Sanskrit-speaking, Indian identity since the third millennium BC.

Whatever their differences, all Indus researchers agree that there is no consensus on the meaning of the script. There are three main problems. First, no firm information is available about its underlying language. Was this an ancestor of Sanskrit or Dravidian, or of some other Indian language family, such as Munda, or was it a language that has disappeared? Linear B was deciphered because the tablets turned out to be in an archaic form of Greek; Mayan glyphs because Mayan languages are still spoken. Second, no names of Indus rulers or personages are known from myths or historical records: no equivalents of Rameses or

Ptolemy, who were known to hieroglyphic decipherers from records of ancient Egypt available in Greek.

Third, there is, as yet, no Indus bilingual inscription comparable to the Rosetta Stone (written in Egyptian and Greek). It is conceivable that such a treasure may exist in Mesopotamia, given its trade links with the Indus civilization. The Mayan decipherment started in 1876 using a sixteenth-century Spanish manuscript that recorded a discussion in colonial Yucatan between a Spanish priest and a Yucatec Mayan-speaking elder about ancient Mayan writing.

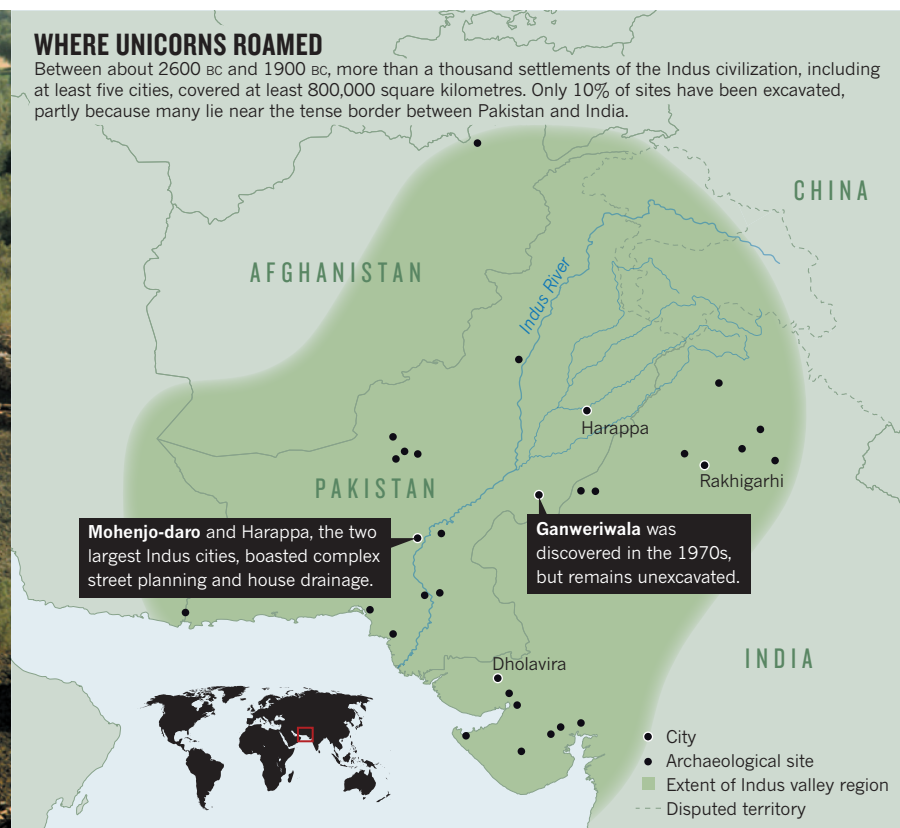
### WHAT WE KNOW

Indus scholars have achieved much in recent decades. A superb three-volume photographic corpus<sup>3</sup> of Indus inscriptions, edited by the indefatigable Asko Parpola, an Indologist at the University of Helsinki, was published between 1987 and 2010 with the support of the United Nations Educational, Scientific and Cultural Organization; a fourth and final volume is still to come. The direction of writing — chiefly right to left — has been established by analysis of the positioning of groups of characters in many differing inscriptions. The segmentation of texts containing repeated sequences of characters, syntactic structures, the numeral system and the measuring system are partly understood.

Views vary on how many signs there are

## WHERE UNICORNS ROAMED

Between about 2600 BC and 1900 BC, more than a thousand settlements of the Indus civilization, including at least five cities, covered at least 800,000 square kilometres. Only 10% of sites have been excavated, partly because many lie near the tense border between Pakistan and India.



from the US-led Harappa Archaeological Research Project.

A group led by computer scientist Rajesh Rao at the University of Washington in Seattle has demonstrated the potential of a digital approach. The team has calculated the conditional entropies — that is, the amount of randomness in the choice of a token (character or word) given a preceding token — in natural-language scripts, such as Sumerian cuneiform and the English alphabet, and in non-linguistic systems, such as the computer programming language Fortran and human DNA. The conditional entropies of the Indus script seem to be most similar to those of Sumerian cuneiform. “Our results increase the probability that the script represents language,” the Rao group has written<sup>9</sup>. Sproat strongly disagrees<sup>10</sup>.

On the ground in Pakistan and India, more inscriptions continue to be discovered — although not, as yet, any texts longer than 26 characters. Unfortunately, less than 10% of the known Indus sites have been excavated. The difficulty — apart from funding — is the politically troubled nature of the region. Many of the most promising unexcavated sites lie in the Pakistani desert region of Cholistan near the tense border with India. One such is the city of Ganweriwala, discovered in the 1970s and apparently comparable in size with Mohenjo-daro and Harappa.

If these sites, and some others within Pakistan and India, were to be excavated, there seems a reasonable prospect of a widely accepted, if incomplete, decipherment of the Indus script. It took more than a century to decipher the less challenging Mayan script, following several false starts, hiatuses and extensive excavation throughout the twentieth century. Indus-script decipherers have been on the much barer trail — older by two millennia — for less than a century, and excavation of Indus sites in Pakistan has stagnated in recent decades. ■

**Andrew Robinson** is a science writer based in London. He is the author of *Lost Languages: The Enigma of the World's Undeciphered Scripts* and, most recently, *The Indus: Lost Civilizations*.  
e-mail: [andrew.robinson33@virgin.net](mailto:andrew.robinson33@virgin.net)

1. Wheeler, M. *The Indus Civilization* 3rd edn 101 (Cambridge Univ. Press, 1968).
2. Fairservis, W. A. *Sci. Am.* **248**, 58–66 (1983).
3. Parpola, A. et al. (eds) *Corpus of Indus Seals and Inscriptions* Vols 1–3.1 (Suomalainen Tiedeakatemia, 1987, 1991, 2010).
4. Rao, S. R. *The Decipherment of the Indus Script* (Asia Publishing, 1982).
5. Parpola, A. *Deciphering the Indus Script* (Cambridge Univ. Press, 1994).
6. Wells, B. K. *The Archaeology and Epigraphy of Indus Writing* (Archaeopress, 2015).
7. Farmer, S., Sproat, R. & Witzel, M. *Electron. J. Vedic Stud.* **11**, 19–57 (2004).
8. Lawler, A. *Science* **306**, 2026–2029 (2004).
9. Rao, R. P. N. et al. *Science* **324**, 1165 (2009).
10. Sproat, R. *Language* **90**, 457–481 (2014).

in the Indus script. In 1982, archaeologist Shikaripura Ranganatha Rao published a Sanskrit-based decipherment with just 62 signs<sup>4</sup>. Parpola put<sup>5</sup> the number at about 425 in 1994 — an estimate supported by the leading Indus script researcher in India, Iravatham Mahadevan. At the other extreme is a high estimate<sup>6</sup> of 676 signs, published this year by archaeologist and epigrapher Bryan Wells.

Nevertheless, almost every researcher accepts that the script contains too many signs to be either an alphabet or a syllabary (in which signs represent syllables), like Linear B. It is probably a logo-syllabic script — such as Sumerian cuneiform or Mayan glyphs — that is, a mixture of hundreds of logographic signs representing words and concepts, such as &, £ and %, and a much smaller subset representing syllables.

As for the language, the balance of evidence favours a proto-Dravidian language, not Sanskrit. Many scholars have proposed plausible Dravidian meanings for a few groups of characters based on Old Tamil, although none of these ‘translations’ has gained universal acceptance.

A minority of researchers query whether the Indus script was capable of expressing a spoken language, mainly because of the brevity of inscriptions. The carvings average five characters per text, and the longest has only 26. In 2004, historian Steve Farmer, computational linguist Richard

Sproat (now a research scientist at Google) and Sanskrit researcher Michael Witzel at Harvard University caused a stir with a joint paper<sup>7</sup> comparing the Indus script with a system of non-phonetic symbols akin to those of medieval European heraldry or the Neolithic Vinča culture from central and southeastern Europe<sup>8</sup>.

This theory seems unlikely, for various reasons. Notably, sequential ordering and an agreed direction of writing are universal features of writing systems. Such rules are not crucial in symbolic systems. Moreover, the Indus civilization must have been well aware through its trade links of how cuneiform functioned as a full writing system.

Nevertheless, the brevity of Indus texts may indeed suggest that it represented only limited aspects of an Indus language. This is true of the earliest, proto-cuneiform, writing on clay tablets from Mesopotamia, around 3300 BC, where the symbols record only calculations with various products (such as barley) and the names of officials.

## DIGITAL APPROACH

The dissident paper has stimulated some fresh approaches. Wells — a vehement believer that the Indus script is a full writing system — working with the geoinformation scientist Andreas Fuls at the Technical University of Berlin, has created the first, publicly available, electronic corpus of Indus texts (see [www.archaeoastronomie.de](http://www.archaeoastronomie.de)). Although not complete, it includes all the texts

**CORRECTION**

In the Comment 'Cracking the Indus script' (A. Robinson *Nature* **526**, 499–501; 2015), the sentence about Bryan Wells's estimate of the number of Indus script signs has been revised to more accurately reflect Wells's contributions. He estimated the number of signs at 676, not 958.