

ON THE DECIPHERMENT OF THE INDUS SCRIPT—
A PRELIMINARY STUDY OF ITS CONNECTION WITH BRĀHMĪ

SUBHASH C. KAK

Department of Electrical & Computer Engineering
Louisiana State University
Baton Rouge, LA 70803
U.S.A.

(Received 10 September 1985)

Arguments are presented that imply that the Brāhmī and the Indus scripts are related. Using similarities of shape between certain Indus signs and the primary Brāhmī signs, the sound values for these Indus signs have been hypothesized. That these sound values are likely to be correct is indicated by the 'decipherment' of a key Indus text.

In the nineteenth century there were two views current regarding the origin of Brāhmī, the script in which most inscriptions of Aśoka are written. According to one, Brāhmī was derived from an older Indian script, whereas the other view saw it as being derived from the Pheonician or South Semitic scripts. In the closing years of the nineteenth century Georg Buhler wrote his influential *Indische Palaeographie* that suggested that Brāhmī signs were originally borrowed from the North Semitic and subsequently modified and enlarged so as to fit the phonetic scheme of the Indo-Aryan languages. Buhler's theory found a lot of favour with those who were convinced that a long silent night existed between the wonderful Harappan civilization (3000 B.C. to 1500 B.C.) and the flowering of the Aryan civilization.

Archaeological evidence unearthed in the past 30 years has pushed the epoch of the coming of Aryans to earlier than 2000 B.C. and therefore the hypothesis that a continuity existed between the Harappan and Aryan civilization becomes very compelling. This hypothesis is further reinforced by arguments powerfully presented by T. Burrow (1973) that the Harappan civilization may itself have been proto-Indoaryan. This makes it imperative that connections between the Indus and Brāhmī scripts be explored carefully.

Some attempts at the decipherment of the Indus script have taken it to be logo-syllabic and then concentrated on the nature of the language of that civilization, which has been taken by these researchers to be Drāvidian or Indo-Aryan with unsatisfactory results in either case. Quite early, Hunter, Langdon and others

argued for a connection between Indus and Brāhmī, yet this line of inquiry has been hampered by a lack of control information. An excellent review of early work that suggests that the language of the inscriptions might be proto-Indoaryan is the book by Mitchiner (1978). For recent reviews of several aspects of this work see Bag (1985), Mahadevan (1985) and Maurer (1985).

The problem of decipherment of the script is compounded by several difficulties. One needs to remember that the texts are from a very long period of several centuries during which many symbols might have changed their forms, even though the script seems to have had great stability. Many of the apparent variants could thus refer to the same symbols. This could apply to differently shaped symbols as well if one were to grant more than one style of the script, which is conceivable given the large geographical extent of the Harappan civilization. Then there exists the problem of the texts being very short generally most being only four or five symbols long. The science of cryptography has shown us that the decipherment of brief texts is, at best, hazardous [see for example Shannon (1949) or the more recent Kak (1980)] because several seemingly valid solutions may be found. In fact for English, if the alphabet is randomly mapped into itself, one needs at least about 27 letters to arrive at a unique decipherment.

This might lead one to give up the problem of decipherment of the Indus script as being utterly hopeless. However, the evidence of continuity between the Harappans and later Aryans indicates that their scripts might be related and therefore there is no need to assume the worst case of random mapping between the two. In fact one may actually find that some characters have not changed very much. Furthermore, it appears that clues from the translations of the texts of the coeval civilization of the Sumerians can provide the control information that allow us to make an advance in the decipherment.

Should it be established that Indus gave rise to Brāhmī, it would leave open the possibility that the Semitic scripts of the Middle East were derived from the Indus script, which should have far-reaching implications for our understanding of the rise of civilizations in the ancient world.

II

Let us briefly state the arguments in favour of the hypothesis that interaction existed between the Harappans and the Aryans :

Aryans and Harappans appear to have coexisted at least for some time as evidenced by the stability of Lothal, and, therefore, there is nothing to require that Aryans could not have lived together with the Harappans, even if as a small minority before the divide of 1500 B.C. Even as an ethnic minority they could have had such influence so as to contribute several words to the language of the

inscriptions. There is considerable archaeological evidence now in favour of the conclusions that the Aryans probably arrived in India before 2000 B.C. [See for example Allchin and Allchin, 1982].

Consider that the Mitannis and Kassites worshipped Vedic deities. In particular, Himalaya is one of the deities in a Kassite record dating back to 1750 B.C. This corroborates archaeological evidence that the Aryans reached India before 1750 B.C., from where some of them went on to the Middle East. This point has been dealt at length by Burrow.

The chronology reconstructed by Pargiter, using the Purāṇas, indicates arrival in India before 2000 B.C., and, therefore, interaction with the Harappans.

Examination of the skeletal remains in Harappan towns suggests mixed populations.

There is a continuity between the Indus and the later Indian weight systems.

Many Indus signs look similar to the Brāhmī signs and this may reflect a real connection. Furthermore, like the Brāhmī Indus script appears to be syllabic, as evidenced by the various forms of the same symbol.

The similarity between one of the Sumerian names for this region : *Bad Imin* [Seven High Places] and *Sapta Saindhava* of the *Rg Veda* and the phonetical similarity between the Sumerian *Dilmun* and Sanskrit *Sindhu* [even though *Dilmun* is commonly believed to have been Bahrein it was influenced by traders from the East, and its weight system corresponds to the one in the Indus region] point to a connection between the Harappans and Aryans, at least in retaining the names of places. Other examples of this are the Sumerian *Meluhha* and *Magan* and the later *Baloch* and *Makran*.

Astronomical allusions in the *Rg Veda etc.* suggests a greater antiquity to the arrival of Aryans in India than is allowed by those who believe that the Aryans had no contact with the Harappans. The analysis of the astronomical evidence has been made by Colebrooke and others, but it has not found general favour. Continuity in the possible Harappan astronomy and the later Aryan astronomy has been discussed by Pargola (1975) and Bag (1985).

The *Rg Veda* and other ancient books do not mention arrival from a outside homeland into *Sapta Saindhava*. If this Veda is dated to 1500-1000 B.C. the Aryans must have lived in this region for centuries to have forgotten their original home.

Indian tradition itself suggests a much earlier point in time as the beginning of their current age.

Recent archaeological findings have also shown that Indus signs were used in several parts of India until about 500 B.C. and sometimes even later.

Without going into the question of the nature of the language of the Harapans, one might ask: It is possible that a connection could exist between Brāhmī, with its 48 letters and Indus script with its more than 300 signs? Many of the indus signs are modifications of a smaller set of basic signs, however. Furthermore, the letters of Brāhmī can be combined together to yield modified symbols.

The most common of these modifications is when a consonant is followed by a vowel and considering all of them would yield about 360 signs. Some of these are not permitted by the syntax of the writing of the Indian languages. Considering the most frequently used modifications of Brāhmī yields a figure that is in the range of 200-300. This is the same order as the number of Indus signs and, therefore, a further investigation into the connection between the two scripts seems justified.

The reasons why the Indus script should be read from right to left are too well known for us to repeat here. However, there are instances where the boustrophedon style was used. Inscriptions in Brāhmī written in the boustrophedon style have also been found, even though Brāhmī was written from left to right. Basham has suggested that Brāhmī might have been originally written from right to left. All this evidence also points to a link between the two scripts.

Having reviewed several arguments supporting possible connection between Brāhmī and Indus, let us summarize the two main objections against this hypothesis.

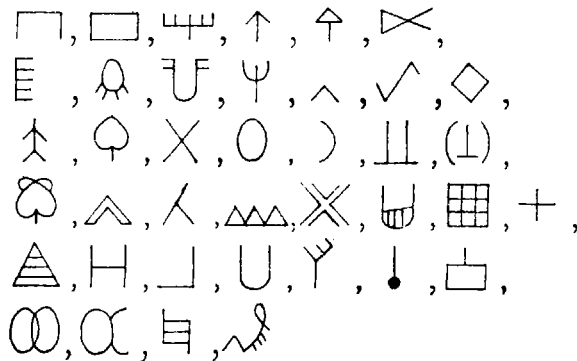
Other known syllabic scripts such as Cretan Linear-B or Elamite cuneiform have less than 200 signs and since Indus has about 400, it is claimed that it could only be logo-syllabic (with both word-signs and syllabic signs). To counter this objection note that the other logo-syllabic scripts of antiquity such as Akkadian, Egyptian and Hittite had between 400 to 900 signs. Considering the possibility that variants of certain signs could have been current in different times, and that some of the signs stand for numerals, the total number of Indus signs does not rule out a syllabic script.

In a logo-syllabic script the word-signs are more in number but occur with low frequency. More than 200 Indus signs occur less than 10 times each. On the other hand, just 2 signs occur over 2000 times out of a total of over 13,000 signs in the corpus. That these figures are not all that much at variance from what one would expect from an alphabetic or syllabic script can be seen by considering English. For a total of 13,000 letters one would encounter *E* about 1700 times and *T* about 1200 times. The least likely letters *J*, *Q*, *X* and *Z* would be encountered about

17, 16, 20 and 10 times each [see for example Konheim (1981)]. The fact that the least likely signs in Indus are so many in number may be due to the texts being mostly names, titles or sacred formulae. In other words the Indus texts are not typical of the Indus written language, whatever it might have been, and thus some departure from typical frequencies is to be expected. It should also be noted that assuming the script to be syllabic does not preclude some symbols to represent concepts when used as abbreviations.

III

Let us examine the similarities between the Indus script and Brāhmī. It has been noted that the 300 odd Indus signs [see for example the Parpola concordance] can be reduced to a small set of primary signs, and the remaining signs may be identified to be combinations of these primary signs. Most of the primary characters can be readily identified; some others are harder to define to be basic. There is further ambiguity from the possible characters to represent numerals. We also consider signs exclusively with vertical strokes to represent numbers, though some exceptions to this may be made. Let us now state our 'first' identification of the primary letters of the Indus script:



These 39 characters, singly or in combinations (with minor 'stylistic' variations) account for more than 80 percent of all the signs. If one were to leave out the signs that are numbers (vertical strokes only) and those that might be higher numbers, such as 10, 20, ...60, etc. one is able to account for more than 90 percent of the total signs. Since our 39 primary Indus signs constitute a set that is of a size comparable to that of the Brāhmī letters (45 signs), the suggestion that there might be a direct correspondence between the two scripts seems to become more credible.

Our initial identification, based essentially on shape (but also other considerations outlined in the next section) is as follows :

Indus	Brāhmī	Sound
H	𑀘	a
L	𑀉	i
⋈	𑀚	u
△	+	ka
⋈	∧	ga
𑀓, 𑀔	𑀛	gha
	𑀜	cha
	𑀝	ja
	𑀞	jha
	𑀟	ṭa
	𑀠	da
∧	𑀡	ta
𑀢	𑀣	da
⊥, 𑀥	⊥	na
	𑀧	pa
□	□	ba
𑀩	𑀪	ma
𑀫	𑀬	ya
),	𑀭	ra
	𑀮	la
□	o	va
↑	↑	sha
𑀱	𑀲	sa
	𑀳	ha

















This provides an identification with 15 of the 24 basic (as determined by Upasak) Brāhmī signs. We consider the secondary Brāhmī signs now :

Indus	Brāhmī	Sound
	π	a
	::	i
	∩	u
	△	e
↗	∧	ai
↘	└┘	o
	└┘π	au
	∩π	am
	∩∩	kha
	∩∩	na
	∩∩	chha
	∩∩	na
	∩∩	ṭha
	∩∩	ḍha
	∩∩	ṇa
	∩∩	ra
○	∩∩	tha
	∩∩	dha
	∩∩	pha
	∩∩	bha
	∩∩	sha

Clearly, identification with the secondary Brāhmī signs has not been equally successful which is further evidence in favour of Brāhmī having been derived from Indus.

Combinations of Indus characters :

We list a random selection.

a-da	
anda	
avada	
shada	
dasha	 or 
shanda(sindh?)	 , 
ru	 , 
runa	
rama	
jja	
kya	
kja	
ko(?)	

This list can be expanded. We caution the reader that we do not claim to have found rules governing the generation of compound signs. The above identifications seem plausible, however.


In summary, it may be said that there seems to be enough justification on the basis of shapes in making the identification of sounds that has been proposed by us. The gaps in the list will have to be filled using contextual cues. Furthermore, it appears that since it was so easy to identify the primary Brāhmī letters amongst the Indus ones, the hypothesis that the former are derived from the latter is likely to be true.



One might argue that the correspondence between the Brāhmī and Indus shapes proposed by us is not unique. An example of an entirely different set of correspondences is the one given by Pandey. Our partial set has been derived using a


comparison of the simplest shapes, and other arguments that will be reviewed in the next section.


IV

To confirm our partial identification, we will examine this problem from a different angle. In particular let us note the similarities with some Sumerian signs as was done by Kinnier Wilson and Mitchiner especially in regard to the identification of *Bad Imin*.

The signs  and variants are similar to the Sumerian signs


,  etc. with the reading *bad* [wall or walled city] and the

Akkadian equivalent *dūru*. *Imin* is a possible reading for , which can be taken to be a compound for *ia* (five) and *min* (two).

The sign  appears a total of 77 times [out of a total of 3204] in the Parpola concordance. A total of 36 times it occurs as the group

$$\begin{array}{c} \text{II} \\ \text{= } \uparrow \text{=} \end{array} \begin{array}{c} \text{IIII} \\ \text{III} \end{array} \quad \dots \quad (1)$$

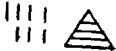


which has therefore been interpreted as *Bad Imin*.

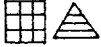


We raise an important question here. Is it possible that the text (1) was read as *Bad Imin* by the Sumerians and *Sapta Saindhabha* by the Harappans? We have already seen in Section III that  can be read as *sha-n-d* which may be an archaic form for *sindhū*, *sinda*, or *sindhava*. The hypothesis that *Bad Imin* represents the Indus area appears to be confirmed, also increasing our confidence in taking Brāhmī to have been derived from the Indus script.

The following variant of *Bad Imin* (or seven *bad*)

$$\begin{array}{c} \text{II} \\ \text{= } \uparrow \text{=} \end{array} \begin{array}{c} \text{IIII} \\ \text{III} \end{array} \quad \dots \quad (2)$$


would actually be read *sapta* (seven) *bad* [□ + □] in the Indus language.









The sign group  also appears to have represented *Bad Imin*. If we begin with the assumption that the region was also called (seven) *grha* then does  somehow represent $\triangle + \text{---} + = =$ *grha*. Of course the Sumerian sound for  is *kur* which is close to that of *grha*.

The inscription  found on copper tablets in Mohenjodaro could then be *grha-rāja* as $\left[\begin{array}{c} \text{E} \\ + \\ ? \\ + \end{array} \right] = ja + a + rā = rāja$, as also identified thus by Mitchiner. This would imply that  as well as  should both represent *ja*.

V

The inscriptions on copper tablets are particularly important because they could represent names of deities, as against accounting records that many seals are likely to have been.

Texts on copper tablets that have been found on more than 20 inscriptions and which end in  are

	
	
	
	fish
	

These would be read

jakya	(yajña ?)
varuna	
thara	(the desert ?)
m-ma-ra-?-ta	(ma-rut ?) or (mitra ?)
m-?-?-o-?	(mitroh ?) or (marut ?)

respectively.

The text



occurs 'nine times on rectangular stamped seals, bearing the motif of a gharial with a fish in its jaws a device which appears on a total of only 39 items in the entire corpus of Indus inscriptions' [Mitchiner, p.46]. This is thus likely to be an important text. Using identification of Section 2, it reads

th(i) j ma-andrama-m-ma-ra-?-ta-varunas

or perhaps

dhvajam (?) indram-mitra-varunas.

The names of these Vedic deities are in consonance with the hypothesis of a continuity between the Harappans and the Aryans.

VI

Assuming that the names of the Vedic gods in certain inscriptions indicate that we are on the right track, we can look at other texts to yield Sanskrit words and Vedic expressions. This could be used to determine the sounds corresponding to other Indus characters. Of course, some of the identifications made by us in Section III could be wrong, but one may place some faith in these at this stage since we were driven by considerations that are objective: matching 'simplest' Indus signs with the corresponding Brāhmī signs.

VII

Our examination of the Harappan signs is continuing. While the names 'deciphered' by us in the previous section occur at several places there is no other significant reading one can point to. There are several reasons for this: (i) we have not found the sounds for all the Harappan characters and therefore many texts are still unreadable, (ii) other texts appear to refer to transactions involving numerical expressions which cannot be used to further our decipherment, (iii) we do not know the language of the inscriptions. Comparing the inscriptions from Mohenjodaro and Harappa that are available to us, we believe that one hypothesis that

cannot be dismissed at this point is that the Indus language could have had at least two variants. This hypothesis would agree with our finding of several variants of *sapta saindhabha*. If this hypothesis were true, one needs to put the symbols in at least two categories [for two variants] that, by making each set smaller, would make further decipherment easier.

The fact that we were able to decipher one of the longest texts is heartening from a cryptologic point of view. We believe that upon revision and expansion of the correspondences between Brāhmī and Indus, further decipherments would follow. Our work does not conclusively prove that Brāhmī and Indus are related, but it makes it highly probable that they are.

BIBLIOGRAPHY

- Allchin, B. and Allchin, R., *The Rise of Civilization in India and Pakistan*, Cambridge University Press, Cambridge, 1982.
- Bag, A. K., *Science and Civilization in India—Harappan Period*, Navrang, New Delhi, 1985.
- Buhler, Georg, *Indische Palaeographie*, 1896. Hindi translation by Mangal Nath Singh, Motilal Banarsidass, Delhi, 1966.
- Burrow, T., The Proto-Indoaryans, *J. Royal Asiatic Soc.*, No. 2, 123-140, 1973.
- Kak, Subhash, The Roots of Science in India, Lecture at the India International Centre, New Delhi, Jan. 3, 1986.
- Kak, Subhash, Masking Ciphers, *Proceedings of IEE*, vol. 127, June 1980, pp. 185-189.
- Kinnier Wilson, J. V., *Indo-Sumerian*, Clarendon Press, Oxford, 1974.
- Konheim, A. G., *Cryptography : A Primer*, John Wiley, New York, 1981.
- Mahadevan, I., S. R. Rao's Decipherment of the Indus Script, *The Indian Historical Review*, 8, Number 1-2, 58-73.
- Mahadevan, I., Claims of Decipherment of the Indus Script, SARC Workshop on Epigraphy, Mysore, 1981-2, March 1985.
- Maurer, W. H., Review of The Decipherment of the Indus Script by S. R. Rao, *J. Amer. Oriental Soc.*, 105.2, 374-378, 1985.
- Mitchiner, J. E., *Studies in the Indus Valley Inscriptions*, Oxford & IBH, New Delhi, 1979.
- Pandey, R. B., *Indian Palaeography*, Motilal Banarsidass, Banaras, 1952.
- Parpola, Asko, Tasks, Methods and Results in the Study of the Indus Script, *J. Royal Asiatic Soc.*, No. 2, 178-209, 1975.
- Rao, S. R., *The Decipherment of the Indus Script*, Asia Publishing House, Bombay, 1982.
- Shannon, C. E., Communication Theory of Secrecy Systems, *Bell System Tech. J.*, 28, 656-715, 1949.
- Upasak, C. S., *The History and Palaeography of Mauryan Brahmi Script*, Nava Nalanda, Nalanda, 1960.
- Vats, M. S., *Excavations of Harappa*, Delhi, 1940.