Indus script

We are writing in response to two articles (Papers 1 and 2) by Srinivasan *et al.*[1] published in *Current Science*.

While the authors make a sincere attempt to understand Indus writing, we believe the conclusions are not supported by the data and the results provided are at variance with different aspects of Indus writing. The articles overlook prior work on entropic and statistical analysis of the Indus script[2–9] that studied many of these issues. We discuss some of our reservations and concerns below, and refer the interested readers to the references for a more detailed discussion (these papers are also available at: www.tifr.res.in/~archaeo).

The entropy values in table 2 of the paper 1 only imply that the first- and second-order entropies of the Indus writing are similar to linguistic writing, a point that was already made by us in ref. 9. As discussed by Rao and co-workers[6,9], this similarity may be a necessary but not sufficient condition to conclude that the writing is linguistic, let alone to conclude that 'the language underlying the Indus text is both Dravidian and Aryan in origin'. The proper way to understand the entropy results is within a Bayesian context, as discussed in Rao *et al.*[9]. Additionally, one needs to look at higher-order entropies (beyond first- and second-order) to gain deeper insights into the structure of the script as explained in Rao *et al.*[9].

It is well known that computing entropy values by only counting frequencies without any smoothing can result in incorrect estimates of the true entropy values of the data. This is because the data is limited and many pairs that can otherwise occur may not have been observed due to incompleteness of the excavated data. Prior work on entropic analysis of the Indus script therefore employed smoothing[3–5] and Bayesian[6,9] techniques. The lack of use of smoothing by Srinivasan *et al.*[1] in their computation of entropy values casts further doubt on the validity of their results.

Apart from these problems of overstretched conclusions from applying information theoretic techniques, there are other problems with the statistics of appearance of various signs and sign combinations that are highlighted in paper 1 against their significance in the corpus. We take some examples of this from paper 1 below. Since paper 2 relies on the results of paper 1 which, in our view, are not tenable, we do not discuss its conclusions here.

(1) The most glaring example is of the authors' use of the sign pair  , which they assign a certain value at several places in the paper and find it significant. In the Indus corpus of Mahadevan[10], this sequence appears only once and that too on an object found in West Asia. The text containing this sequence is, according to the likelihood calculations, an extremely improbable sequence[4].

(2) There are disparities in many of the linguistic associations of the Indus signs with various elements of different languages. The frequency of appearance of a large fraction of the identified associations is often so rare as to be insignificant. For example, the 'initial vowel'  (paper 1, p. 154) appears only once in the corpus.

(3) The authors state that 'An alternative form to indicate the presence of long medial-vowels "ē" () and "ō" () and diphthong "ai" () exists in the Indus texts' (paper 1, p. 154). These signs in fact have a positional sensitivity with the signs moving increasingly towards the middle of the texts from the text-ending position as the number of strokes increases[7]. This is at odds with their phonetic association in normal language structures.

(4) Other problems include the suggestion that the signs pairs  and  represent 'Aryans' and 'Dravidians' (paper 1, p. 156). The pair appears nine times (eight times in Mohenjo Daro and once in Harappa, and never at any other site), while the sign sequence appears only once in West Asia. The statistics therefore casts serious doubt on the claim of the authors.

(5) The list of points in the section 'Some important observations' (paper 1, p. 156) does not seem to have any significance. With 417 signs, one can list out 417 × 417 (173,889) such sign combinations which are possible, out of which only about 2758 distinct pairs appear in the dataset. Hence there are more than 171,131 sign combinations which do not appear and it is not at all clear why non-occurrence of only some pairs is considered important.

(6) The comparison with Brahmi is another issue (paper 1, p. 154), whose statistics are not given in table 2. The associations of Brahmi characters with the Indus signs and sign sequences are made without any basis. Incidentally such comparisons have been tried by other scholars earlier (see Possehl[11] for a critical review of such attempts) and are not considered beyond chance.

Although the authors[1] have made efforts to compare the Indus writing with other writing systems, the problem of its nature is not resolved. Also, our recent work on design of the Indus signs[7] has shown that the signs which appear to be a conglomeration of individual signs are in fact far more complex than made out by the authors.

The question of the structure and nature of the Indus script therefore remains as enigmatic as ever. To summarize, the conclusions of paper 1 fail on the following grounds:

(1) The claims of similarity based on entropy arguments are not conclusive since similarity of first- and second-order entropies can arise due to several reasons even between unrelated systems. The articles assume that the bigram similarity is enough to make detailed comparison of the different linguistic systems.

(2) Even with this assumption, Paper 1 makes other claims of similarity which are not tenable. While assigning the values to signs the authors do not take into account the position sensitivity of these signs and sign sequences which invalidates several of their conclusions (Yadav *et al.*[2]).

(3) The authors have made major claims based on signs and sign pairs that are extremely rare and appear in unusual conditions[4,5].

(4) The claims made by the authors on the design of signs are inconsistent with several aspects of their design and environment[7].

In summary, we believe that the articles have serious shortcomings and fail in their claim that the Indus script is of abugida type and underpins current Indian linguistic scripts.

1. Srinivasan, S., Joseph, J. V. M. and Harikumar, P., *Curr. Sci.*, 2012, **103**, 147–157; 268–281.

2. Yadav, N., Vahia, M. N., Mahadevan, I. and Joglekar, H., *Int. J. Dravidian Linguist*., 2008, **XXXVII**, 39–52; 53–72.
3. Rao, R. P. N., Yadav, N., Vahia, M. N., Joglekar, H., Adhikari, R. and Mahadevan, I., *Science*, 2009, **324**, 1165; http://www.cs.washington.edu/homes/rao/ScienceIndus.pdf
4. Rao, R. P. N., Yadav, N., Vahia, M. N., Joglekar, H., Adhikari, R. and Mahadevan, I., *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 13685–13690.
5. Yadav, N., Joglekar, H., Rao, R. P. N., Vahia, M. N., Adhikari, R. and Mahadevan, I., Statistical Analysis of the Indus Script Using n-Grams. *PLoS ONE*, 2010, **5**, e9506; doi:10.1371/journal.pone.0009506.
6. Rao, R. P. N., Yadav, N., Vahia, M. N., Joglekar, H., Adhikari, R. and Mahadevan, I., *Comput. Linguist.*, 2010, **36**, 795–805.
7. Yadav, N. and Vahia, M. N., *Scripta*, 2011, **3**, 133–172.
8. Yadav, N., *Man Environ.*, 2012, **XXXVII**, 1–7.
9. Rao, R. P. N., *IEEE Comput.*, 2010, **43**, 76–80; doi: 10.1109/MC.2010.112.
10. Mahadevan, I., *The Indus Script: Texts, Concordance and Tables*, Memoirs of the Archaeological Survey of India No. 77, New Delhi, 1977.
11. Possehl, G. L., *Indus Age: The Writing System*, Oxford IBH Publishing Co. Pvt Ltd, New Delhi, 1996.

NISHA YADAV[1,*]
RAJESH P. N. RAO[2]
M. N. VAHIA[1]

[1]*Department of Astronomy and Astrophysics, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400 005, India*
[2]*Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195, USA*
**e-mail: y_nisha@tifr.res.in*

Response

We appreciate the comments made by Yadav *et al.* on our articles. They have expressed disagreement with our conclusion that the Indus script is abugida type. However, we believe that the statistics given in tables 3 and 7 and some important observations (1–4) listed in paper 1 (p. 156) do support the presence of medial vowel signs in Indus script and their semblance to Dravidian-type language such as Kannada. The findings described in paper 2 are of importance and cannot be ignored. The six predominantly used fish-like symbols present in the Indus writing bear one-to-one semblance with the akshara symbols for va, pa, ma, ba, sha, gha present in the modern Kannada script. Further, two anticipated labial aksharas for 'pha' and 'bha' that were missing in IM corpus were found from BW corpus. Besides that, Indus signs for fractional numbers (1/4, 1/2, 3/4, 1/16, 1/8, 3/16) were identified in the Indus text. We strongly feel that the observations made in our articles are deep-rooted from a linguistic point of view and would certainly lead a step forward in understanding the genesis of modern Indian language scripts.

S. SRINIVASAN[1,*]
J. V. M. JOSEPH[2]
P. HARIKUMAR[3]

[1]*Apartment #43, First Avenue,*
[2]*Apartment #98, Fifth Avenue, Pudupattinam Colony, Kalpakkam 603 102, India*
[3]*Plot No. 202, Fifth Street, Rajeswari Nagar, Kelambakkam 603 103, India*
**e-mail: indussrini@gmail.com*

Mutation in impact factor

Acta Crystallographica Section A: Foundations of Crystallography [pISSN 0108-7673] exhibited perhaps the first ever mutation in the impact factor (IF) of a journal. Specifically, its IF was merely 2.051 in 2008, mutated to 49.926 in 2009 and even increased to 54.333 in 2010; now its 2011 IF is 2.076, i.e. what it should have been[1].

The main reason attributed to this high IF was a single paper authored by George Sheldrick, 'A short history of *SHELX*' (2008, **64A**, 112–122) providing an account of the development of the *SHELX* system of computer programs from 1976 to date[2].

This brings a few points to the fore, i.e. a review paper plays an important role in determining the IF of a journal. Also, one must look at the citation of an individual paper while undertaking any evaluation exercise, as IF alone may not provide the correct picture[3].

More importantly, isolated cases like this have prompted the use of *h*-index for evaluation exercise, especially for senior scientists[4]. And even *p*-index[5] is knocking!

1. *Journal Citation Reports 2009, 2010, 2011, 2012*, science edition (web-based), Thomson Reuters, Philadelphia, 2009, 2010, 2011, 2012.

2. Jain, N. C., *Ann. Libr. Inf. Stud*., 2011, **58**, 87.
3. Jain, N. C., *Curr. Sci*., 2011, **101**, 717.
4. Hirsch, J. E., arXiv:physics/0508025v5 [physics.soc-ph] 29 Sep 2005.
5. *p*-index, available from http://giridhar-madras.blogspot.in/2009/07/p-index.html, accessed on 13 September 2012.

N. C. JAIN

Division of Publication and Information, Indian Council of Medical Research, V. Ramalingaswami Bhawan, Ansari Nagar, New Delhi 110 029, India
e-mail: drencejain@gmail.com