ARTICLE

<https://doi.org/10.1057/s41599-021-00713-0>

OPEN

A method of identifying allographs in undeciphered scripts and its application to the Indus Valley Script

Shruti Daggumati¹ & Peter Z. Revesz¹

This work describes a general method of testing for redundancies in the sign lists of ancient scripts by data mining the positions of the signs within the inscriptions. The redundant signs are allographs of the same grapheme. The method is applied to the undeciphered Indus Valley Script, which stands out from other ancient scripts by having a large proposed sign list that contains dozens of asymmetric signs that have mirrored pairs. By a statistical analysis of mirrored asymmetric signs, this paper shows that the Indus Valley Script was multi-directional and the mirroring of signs often denotes only the direction of writing without any difference in meaning. For this and five other specific reasons listed in the paper, 50 pairs of signs, 23 mirrored, and 27 non-mirrored, can be grouped together because each pair consists of only insignificant variations of the same original sign. The reduced sign list may make decipherment easier in the future.

¹Department of Computer Science and Engineering, University of Nebraska-Lincoln, Lincoln, USA. ✉email: revesz@cse.unl.edu

Introduction

An important first step in understanding a script is being able to tell how many different signs it has. Hence variations of signs that denote the same concept need to be grouped together. Although this grouping is naturally present in all work, the process of grouping is not described and analyzed in a systematic way but mostly presented as a fact based on the visual sense of the researcher with only occasional ad-hoc arguments given. Our goal is to systemize the process of grouping of variations of signs. The debate over the exact number of signs is particularly acute in the case of the Indus Valley Script. The Indus Valley Script is estimated to have between 417 (Mahadevan, 1977) and 694 (Wells, 2015) signs. The Indus Valley Script also shows a large variation in the way certain signs are written because of the various ways it was expressed including carving, chiseling, embossing, incising, inlaying, molding and painting, and the range of writing materials used including bone or ivory, ceramic, copper, faience, gold, gypsum, sandstone, silver, steatite, stoneware, and terracotta (Parpola et al., 2010). When consulting the largest collection of signs (Wells, 2015), <10% occur more than 50 times as shown in Fig. 1.

It is also necessary to establish the correct reading direction of each inscription. Previous scholars who have made over a hundred unsuccessful attempts to decipher the Indus Valley Script have mostly concurred that the Indus Valley Script was always read right to left. In addition, some scholars have also accepted that each sign variation is unique. We show that the Indus Valley Script was multi-directional and has a smaller sign set.

In this work, we focus on the asymmetric signs as well as similar signs. We define symmetry as vertical line symmetry, where the left and right portions of a sign are mirror images of each other. It is seen that even though there is an almost equal distribution of asymmetric and symmetric signs, the frequency of symmetric signs is much greater than asymmetric (Daniels and Bright, 1996; Mahadevan, 1977).

When focusing on the asymmetric signs, a couple of interesting characteristics appear; duplicate and mirrored signs. Mahadevan

and many other scholars have neglected the difference between the original and reversed sign on seals. Wells is one scholar who noted the difference between the original and mirrored signs and regards them as separate entities (Wells, 1998). Therefore, it is essential to correctly group signs based not only on their visual similarities, as earlier authors have done but also by careful data analysis of their positions within the inscriptions (Fuls, 2013; Wells, 2011, 2015).

The rest of this paper is organized as follows. Section “Data sources and methods” describes the data sources used. Section “Multi-directionality of the Indus Valley Script shown by mirrored asymmetric signs” considers asymmetric signs that have mirrored pairs. Section “Reducing the Indus Valley Script sign set by data mining inscriptions” discusses the implications of merging some groups of signs and thereby reducing the total number of signs in the Indus Valley Script. Section “Conclusions and future works” gives some conclusions and presents ideas for future work. Finally, section “Data availability” makes a statement about data availability.

Data sources and methods

The signs noted in this work reference multiple authors (Mahadevan, 1977; Parpola, 1986, 1994; Wells, 1998), CISI (Joshi and Parpola, 1987; Parpola et al., 2010; Shah and Parpola, 1991), and the ICIT dataset. The data set we focus on was curated and verified in two ways. First, by hand (using sign lists from other authors and CISI), and second, by using the ICIT database as a resource. Each sign on the seals in question were stored in a MongoDB database. The signs we focused on for symmetric/asymmetric were a primary field that enabled us to focus on their relation to other signs in the seal and similar seals. The following attributes were stored for each seal: CISI id, sign number, location, other signs on the seal, length of the seal, and a flag to indicate if it was a multi-line seal. Each seal is stored as a document which has the aforementioned properties. Unlike traditional databases the MongoDB database allows for multiple correlations to be made to a sign and it also allows for an easier analysis. Each of the frequencies listed in this work is easily tabulated via querying the data set. This database setup could expand in the future to further analyze seals with animal symbols.

Multi-directionality of the Indus Valley Script shown by mirrored asymmetric signs

In this section, we focus on the signs which we classify as mirrored. According to our identification, there are at least 23 mirrored asymmetric signs in the Indus Valley Script. Each of the mirrored asymmetric signs can be used multiple times, but usually much less frequently than the original sign. Hence while the original signs occurred 1659 times, their mirrored signs occurred only a total of 110 times, that is only about 6.7 percent. As we went through each example of a mirrored sign, we tried to find an apparent reason for the mirroring. We divide the occurrence of mirrored signs into two distinct categories: (I) deliberate and (II) accidental. We could distinguish five different reasons for a deliberate mirroring of a sign. When we could not find an apparent reason, then we call the use of the mirrored sign ‘accidental.’ In general, we tried hard to find a reason for mirroring of the signs and accepted that a mirroring is accidental only if no other reasonable explanation could be identified. As we browsed through the library, we found that the earlier classifications of the reasons for the mirroring were useful in identifying the later occurrences of mirrored signs. Each category has various types, which are listed below:

Fig. 1 Distribution of Indus Valley Script signs. Venn diagram regarding asymmetry, mirroring, and frequency.

Classification of mirrored signs

I - Deliberate

Type 1. The mirrored sign is found in a sequence that is reversed.

Type 2. Two or more mirrored signs occur on the same seal/tablet.

Type 3. The mirrored sign appears on a crowded seal to save space.

Type 4. The mirrored sign occurs on a boustrophedonic artifact.

Type 5. The mirrored sign indicates an underlying meaning.

II - Accidental

Type 6. The mirrored sign is incorrect in the ICIT database of (Wells and Fuls, 2017).

Type 7. The mirrored sign is a location anomaly.

Type 8. The mirrored signs (a) are carving errors or (b) occur at different periods.

Type 1 occurs when all asymmetric signs are mirrored. This may indicate that instead of the normal right to left reading direction a reverse direction is intended.

Type 2 also may indicate a left to right reading direction.

Type 3 is merely to save space.

Type 4 reverses the reading direction for each row to make reading possible without a large movement of the eyes.

Type 8 is divided into cases. If an anomaly occurs only once, then it is carving error (a), while if it occurs several times but always in late periods, then it is a style change over time (b). We consider these two cases related in the sense that an accidentally introduced carving error could be copied by later scribes.

Neither Types 1–4 nor Types 6–8, the accidentally mirrored occurrences, have grammatical meanings. Only Type 5 allows the mirroring to denote a grammatical marker. We categorized as Type 5 only some of the occurrences of the mirrored pair 𑀩 and 𑀪. Wells (2011) has done a detailed statistical analysis of some of these symbols and argued for a grammatical marker role for the mirrored signs. We feel that Types 1–4 and 6–8 occurrences should not be used in any statistical analysis regarding grammatical meaning because of the non-grammatical explanations.

Data collection. There are 23 pairs of mirrored signs. The curated sign list is retrieved from the up-to-date online Indus writing database which hosts the Interactive Corpus of Indus Text (ICIT) (Wells and Fuls, 2017). In addition to using ICIT, we referenced the Corpus of Indus Seals and Inscriptions (CISI) to verify the mirrored signs (Joshi and Parpola, 1987); in our work, we only analyze the signs found in CISI.

Table 1 shows the 23 pairs of mirrored signs, where the left sign is the original sign and the right sign is the mirrored sign. Wells (2015) already noted that the signs, 𑀢, 𑀣, 𑀤 and 𑀥 have the four largest allographic sets of Indus Valley Script signs. We claim that all the 23 pairs of mirrored signs are allographs. The frequencies listed next to each sign are also retrieved from ICIT. As seen in the table each mirrored sign occurs <20 times, where almost half occur at most twice. Table 1 also contains the artifact names where the mirrored signs occur and its' sub-type. Later sections describe the categorizations in detail.

Signs

Sign 𑀩. There are five occurrences of the mirrored sign 𑀪. Only one occurs in Harappa, on tablet H-795 that is barely legible. Assuming that this is sign 𑀩, we state that the tablet is mirrored

due to sign 𑀩 occurring as a terminal character on the left; for the majority of the tablets sign 𑀩 occurs on the right. Seal K-45 occurs in Kalibangan; Kalibangan has only two noted seals with the original sign 𑀩. Due to the scarcity in the region, the evidence shows that it is a location anomaly.

There are three occurrences of the mirrored sign in Mohenjo-daro. Seal M-1783 is boustrophedonic as the mirrored sign 𑀪 appears on the second line. The claim that the second line is boustrophedonic is not supported by the positional analysis work of Fuls (2013), however the original 𑀩 occurs one time on a two-line seal as a terminal character, much like seal M-1783. It was also seen that the signs are not carved as neatly as other Indus Valley Script seals. M-2027 is a fragment of a tablet that is merely sketched in CISI; therefore, we cannot guarantee its validity. The evidence shows that Seal M-632 is carved incorrectly due to its appearing with two other asymmetric signs and due to that the probability of a mirrored sign with the aforementioned classifications is 0.04%.

Sign 𑀢. Four occurrences exist for the mirrored sign. Of those four, entry H-1468 is found on a pot shard, so the direction is unclear; therefore, we conclude that direction is the same as the original. Seal M-62 contains the mirrored sign paired with the original sign with a sign in between; we conclude that this sign is paired, i.e. 𑀢𑀣𑀢.

The last two occurrences are found in Kalibangan. Seal K-6 is a boustrophedonic inscription that contains two mirrored signs: 𑀢 and 𑀣. In addition, the probability that these two mirrored signs occur on the same seal is $\frac{4}{17,650} \times \frac{3}{17,650} = 3.852 \times 10^{-8}$. One interesting aspect of this seal is that the direction that the animal is facing is left when they are most often facing right. Seal K-56 is noted incorrectly in the ICIT database, and the direction should be that of the original.

Sign 𑀤. Three occurrences exist for the mirrored sign. One in Kalibangan and two in Mohenjo-daro. In Kalibangan, Seal K-6 notes the only time either the original or mirrored value of this sign is used. As discussed above, Seal K-6 is boustrophedonic.

The other two seals are found in Mohenjo-daro. Seal M-747 is written in a boustrophedonic manner and should be read from right to left, and then left to right. Seal M-782 is noted incorrectly in the ICIT database and the direction should be that of the original.

Sign 𑀦. According to ICIT, six occurrences exist for 𑀦, three in Harappa and three in Mohenjo-daro. Figure 2 shows two instances of using the 𑀦 sign instead of the 𑀧 sign. The seals shown on the left and the right of this figure are numbered H-50 and M-1350, respectively, in CISI. In the first case the sign is used to save space, while in the second case the sign may be a carving error.

According to ICIT, Tablet M-600 includes mirrored sign 𑀦 on the front side and mirrored sign 𑀧 on the backside. However, on further inspection, we see that the backside has the sign 𑀨. The fragment of the tablet reads on the front 𑀦𑀧𑀦 and 𑀨 on the back. Copper Tablet M-599 contains the same text on the front and back as M-600. Given the evidence of two mirrored signs on the front and back the evidence shows that the seal contains mirrored writing and that the probability of these two mirrored signs occurring on the same seal is $\frac{8}{17,650} \times \frac{3}{17,650} = 7.704 \times 10^{-8}$. These tablets are similar to copper tablets M-547, M-548, M-549, M-1555, M-1556, M-1557, and M-1558, where the only difference is the mirrored sign. Upon inspecting seal M-749 in CISI we see that it does not contain sign 𑀦 but instead 𑀧.

For the three occurrences in Harappa, the signs found on the pottery artifacts H-1413 and H-1745 very slightly represent sign

Table 1 Signs which are found to be mirrored, including their frequency, CISI ID, and categorized type.

Original		Mirrored		Artifact	
Sign	Frequency	Sign	Frequency	CISI ID(s)	Type
𐀀	54	𐀁	5	H-795	1
				M-1783	4
				K-45	8
				M-632	
				M-2027	
𐀂	113	𐀃	4	M-62	6
				K-56	8
				H-1468	
				K-6	2
𐀄	20	𐀅	3	M-747	4
				M-782	6
				M-749	6
𐀆	121	𐀇	6	H-1413	8
				H-1745	
				H-1930	
				M-599	2
				M-600	6
𐀈	135	𐀉	4	H-362	
				B-4	7
				H-2104	8
𐀊	8	𐀋	1	H-1775	8
𐀌	455	𐀍	3	M-1233	3, 7
				K-77	8
				L-251	
𐀎	133	𐀏	15	H-8	3
				H-1657	
				H-2128	
				M-239	4
				M-1272	6
				H-209	8
				H-482	
				M-494	
				H-64	5
				H-161	
				M-245	
				M-779	
				M-1141	
				M-1737	
				M-1826	
𐀐	67	𐀑	17	H-278	3
				H-279	
				H-280	
				H-281	
				H-282	
				H-283	
				H-284	6
				H-598	
				H-1924	
				M-189	7
				H-834	8
				H-871	
				H-872	
				H-873	
				H-1154	
				H-1835	
				H-2039	
𐀒	17	𐀓	2	M-331	4
				Ch-4	8
𐀔	1	𐀕	2	H-1902	8
				M-402	4
𐀖	2	𐀗	2	K-18	6
				M-745	
𐀙	7	𐀚	3	L-90	6
				M-525	
				M-1322	
𐀛	87	𐀜	4	H-305	6
				H-2124	8
				H-1864	
				M-664	
𐀞	192	𐀟	6	H-232	6
				H-1302	
				H-1303	
				H-1822	
				M-1650	
				H-853	8
𐀠	107	𐀡	12	H-50	3
				H-663	6
				L-28	7
				L-37	
				L-210	
				H-1409	8
				M-170	
				M-848	
				M-1047	
				M-1350	
𐀣	7	𐀤	1	M-1961	
				M-2061	
𐀥	11	𐀦	1	L-11	7
𐀨	63	𐀩	10	M-293	7
				H-344	8
				H-359	
				H-715	
				H-833	
				H-968	
				H-2084	
				H-2142	
				L-9	
				M-63	
𐀬	3	𐀭	2	M-934	8
𐀮	50	𐀯	2	H-1289	
				M-784	8
𐀱	2	𐀲	2	H-1920	
				H-298	8
𐀴	4	𐀵	3	M-226	
				M-1690	8
				M-702	
𐀷	4	𐀸	3	M-976	8
				M-1688

Fig. 2 Mirroring of the rake sign. Seal 131 (left) from Vats (1940) and seal 599 (right) from Mackay (1938) show the use of the 𑀩 sign.

𑀩. Triangular Tablet H-1930 contains a few peculiarities. The sign 𑀩 does not have the “E” shape that 𑀪 has and more of a “Y” shape. In addition, this is the only tablet to have both signs 𑀪 and 𑀫. With all of these unique attributes, the evidence supports a negligible mirroring.

Sign 𑀬. Four occurrences exist for mirrored sign 𑀬. Mohenjo-daro has one occurrence of the mirrored sign on the copper Tablet M-600; as discussed above.

One mirrored sign is found on Seal B-4 in Banawali which reads 𑀬𑀭. In this location, sign 𑀬 only occurs once, and that is in the mirrored format. Due to its rarity in this location; it is a location anomaly.

In Harappa, we find the other two occurrences of 𑀬. Tablet H-2104 is unique where the sign 𑀬 appears only twice throughout the Indus Valley Script. Upon consulting CISI we see that the direction noted in ICIT is incorrect and the sign, in reality, is the original. Triangular Tablet H-632 is unique in regards to its shape and reads 𑀬𑀭, where we see that the terminal sign is 𑀭; therefore, we can conclude that this tablet likely has an erroneous carving.

Sign 𑀮. The ICIT database contains one account of the mirrored sign 𑀮 which is found on the H-1775 tablet as listed in CISI. Given that this is one account of this sign, and it contains a common triplet pair we conclude that this seal contains an accidental mirrored carving.

Sign 𑀯. Out of 455 occurrences of sign 𑀯 noted in ICIT only three occur as a mirror 𑀰. This low percentage is our first inclination that the mirrored sign does not indicate a different glyph. The mirrored sign that occurs independently on a shard of pottery L-251, allows us to conclude that directionality of the sign is the original. Seal K-77 reads 𑀯𑀰 which is similar to Seal K-68 which reads 𑀰𑀯 only mirrored. We see that the obverse side of the seal has similar signs i.e. long vertical lines. We also see that the seals are similar in shape and size.

Consulting CISI we see that Seal M-1233 has a crude unfinished surface with a sign that appears to be 𑀰 with one less horizontal line. In addition, we notice that sign 𑀰 is crowded at the edge of the seal, and the carvings seem more haphazard than usual Indus Valley Script carvings. We conclude that this single occurrence in Mohenjo-daro is likely an erroneous carving.

Sign 𑀱. The mirrored sign 𑀱 appears 15 times according to the ICIT database. Figure 3 shows two examples of the use of the 𑀱 sign. The seals shown on the left and the right of this figure are

numbered H-8 and H-161, respectively, in CISI. Wells (2011) has performed a detailed analysis on this symbol. We see the most unique occurrence of the mirrored sign on Seal M-1272; where ICIT states that the mirrored sign appears with the original. Inspecting the seal in CISI we see that the stated mirrored sign appears on the right-hand side, as shown on seal M-1272. Upon further inspection we note that the first sign is different from the one stated, and the curved nature of the first sign is not similar to the stated mirrored sign (it is not truly symmetric). The quoted mirrored sign is more similar to 𑀲, which occurs as an initial symbol all but one time throughout all of the Indus Valley Script seals.

In seals H-8, shown on the left side of Fig. 3, and H-1657 the mirrored sign appears as a pair with sign 𑀳, where the right-hand side is crowded with 𑀳. We see that the 𑀱 fits perfectly into the angles of the 𑀳 sign. An overlap of artifact sequences among the items where the original signs and mirrored sign are found is non-existent. Upon further inspection, it appears that the 𑀱 sign appears in the sequence with either 𑀴𑀵 or just 𑀴, whereas the combination of 𑀴𑀶 or 𑀶 is non-existent in the artifacts where the original sign is found. These sequences are found on artifacts H-64, H-161 (see the right side of Fig. 3), H-245, M-245, M-779, M-1141, M-1737, and M-1826. In these cases the mirrored sign may indicate some meaning.

Tablet H-209 is stated to have sign 𑀷, but on further inspection, we see that the sign is not clear. Seal H-482 is stated to have sign 𑀷 on the seal, but with further inspection of CISI, we see that the sign is incomplete and may not be the mirrored sign. Inspecting tablet M-494 in CISI it is not apparent that the sign 𑀷 is on the artifact. Tablet H-2128 is crowded on the left side where the mirrored seal appears; therefore, we conclude that the sign direction is negligible. Seal M-239 is noted to be mirrored as the sign 𑀸 is also mirrored (some scholars may perceive the seal to include the sign 𑀸 instead which has only one occurrence in the entire Indus seal set).

Sign 𑀹. Of the 17 mirrored occurrences listed in ICIT, 11 of them (H-278–H-284, H-871–H-873) are a repeated sign sequence found on a cylindrical clay tablet with a crocodile on the back. We note that the original sign 𑀹 does not occur in this sequence nor does the original sign occur with a crocodile or on a cylindrical tablet. We conclude that these 11 mirrored signs are one repeated mistake.

Seal H-598 and Tablet H-1924 are both incorrectly noted in ICIT, the sign should be noted as the original. H-1154 and H-1835 appear on small tablet shards where the sign is unclear; therefore, the sign may not be 𑀹. Tablet H-2039 has 𑀹 rotated 90°,

Fig. 3 Mirroring of the bow sign. Seals 1 (left) and 262 from Vats (1940) show the use of the $\hookleftarrow$ sign.

thus the evidence supports a writing anomaly. H-2039 appears on a coin size tablet where the sign $\hookleftarrow$ appears squished due to space limitations; thus, the mirroring is negligible.

All but one-mirrored sign occur in Harappa. Seal M-189 is the only one to occur in Mohenjo-daro; therefore, the evidence supports a location anomaly.

Signs $\hookleftarrow$ and $\hookrightarrow$. According to ICIT, there are two entries for $\hookleftarrow$, however, after consulting CISI, we conclude that neither are mirrored signs. One entry is unclear as it is found on a broken shard of pottery; incomplete and may not be the sign stated. The second entry is in good condition; however, the seal M-331, is boustrophedonic. The sign is found in the second row of the seal and the first row is read from right to left.

Seal H-1902 we note to be an accidental carving, due to all other occurrences of this sign are in the initial position. Similar to $\hookleftarrow$ we conclude entry M-402 for $\hookrightarrow$ is also boustrophedonic. The seal has one asymmetric sign which only occurs once throughout the Indus Valley Script which we believe to be a simpler version of $\hookrightarrow$.

Signs $\hookleftarrow$ and $\hookrightarrow$ for the majority of artifacts, are initial characters, thus this ensures that these signs are not mirrors yet, boustrophedonic. In addition, to concluding that signs $\hookleftarrow$ and $\hookrightarrow$ are the originals. We speculate that sign $\hookrightarrow$ is a variation of sign $\hookleftarrow$, which has only occurred once throughout the Indus Valley Script. Supplementary to this we see that neither signs $\hookleftarrow$ and $\hookrightarrow$ nor their mirrors occur together.

Sign $\hookrightarrow$. The ICIT database states two seals contain the mirrored sign $\hookrightarrow$. After further investigation of CISI, we see that K-18 is incorrectly noted and does not contain the mirrored sign, but rather the original. According to ICIT the seal found in Mohenjo-Daro M-745 reads $\hookrightarrow \hookrightarrow \hookrightarrow \hookrightarrow \hookrightarrow$. Upon consulting CISI, we see that this sign is not actually mirrored but the original sign.

Sign $\hookleftarrow$. We note that this sign is among other sign variants as noted by Parpola (1994) and Wells (2015). According to ICIT we see that there are three occurrences for the mirrored sign and seven occurrences for the original sign. However on further inspection of the original sign and the mirrored sign, the artifacts in CISI show only one occurrence of $\hookleftarrow$ and the other nine are the sign $\hookrightarrow$. All but three artifacts are found on copper tablets. For the single seal L-36 which contains the sign $\hookleftarrow$, it is seen that the ticks at the top of the sign would be cut off if the sign were $\hookrightarrow$; therefore, this is due to space limitations on the seal. A mirrored sign does not exist and the only sign is $\hookleftarrow$.

Table 2 Three of the CISI tablets which are most similar to H-1864.	
CISI of artifact	Signs
H-1864	$\hookleftarrow$ $\hookrightarrow$ $\hookleftarrow$ $\hookrightarrow$ $\hookleftarrow$ $\hookrightarrow$
M-164	$\hookleftarrow$ $\hookrightarrow$ $\hookleftarrow$ $\hookrightarrow$ $\hookleftarrow$ $\hookrightarrow$
M-717	$\hookleftarrow$ $\hookrightarrow$ $\hookleftarrow$ $\hookrightarrow$ $\hookleftarrow$ $\hookrightarrow$

Sign $\hookleftarrow$. The mirrored sign occurs four times in the ICIT database. We find that one of the entries, H-305 is noted incorrectly in the ICIT database and should be similar to tablets H-921 or H-2123 which both read $\hookleftarrow \hookleftarrow$. We note that Tablet H-1864 is incorrectly carved, due to similarity to other seals, e.g. M-164 and M-717; illustrated in Table 2. The mirrored sign on tablet H-2124 occurs independently. Therefore, nothing can be concluded other than the sign is a writing anomaly on a tablet, which is in poor condition.

Seal M-664 found in Mohenjo-Daro is incorrectly carved due to that the mirrored sign $\hookleftarrow$ is the only occurrence in this location. In addition, the pair $\hookleftarrow \hookrightarrow$ occurs four times in Mohenjo-daro, as the beginning signs of seal M-664 with sign $\hookleftarrow$ reversed. In addition, the carving on the tablets is slightly written on a diagonal. Signs are usually carved straight across the top of the seal. We see that with the two erroneous mirrored signs the probability of a mirrored sign is 0.02%.

Sign $\hookrightarrow$. The mirrored sign occurs five times in Harappa and once in Mohenjo-Daro. In Harappa, the signs $\hookleftarrow \hookrightarrow$ occur together in this order at least 42 times. Examining tablet H-232, we see that the tablet reads $\hookleftarrow \hookrightarrow \hookleftarrow \hookrightarrow$ where the mirrored sign is not $\hookleftarrow$ instead it is $\hookrightarrow$. In addition, we see that the reason behind mirroring the sign $\hookleftarrow$ is due to space limitations on the tablet. Tablets H-1302, H-1303, and H-1822 are all mirrored writing which contain the signs $\hookleftarrow \hookrightarrow$ however, it is incorrectly noted in the ICIT database as $\hookleftarrow \hookrightarrow$ where the mirrored sign $\hookleftarrow$ should be noted as $\hookrightarrow$. The mirrored sign found in Mohenjo-Daro appears on ivory rod M-1650. On further inspection of CISI, we see that the mirrored sign is also incorrectly noted in the ICIT.

Triangular Tablet H-853 is the only occurrence of the mirrored sign $\hookleftarrow$, as we have concluded the other noted mirrors are incorrect. We see that the probability of having a mirrored sign instead of the original as 0.005%. As the sign is also carved in the unique triangular tablet, we conclude that the tablet is an accidental carving.

Sign. There are 12 occurrences of the mirrored sign . Seal H-663 is noted incorrectly in the ICIT database and should be written as the original. All the occurrences of the mirrored sign in Lothal (L-28, L-37, L-210) are location anomalies, as the original sign and the mirrored sign have an almost equal distribution. For the rest of the artifacts, we see that there are portions of the text inclusive of the mirrored sign which is found on artifacts with the original sign. This evidence shows that mirroring is negligible.

Location anomalies. Of the 23 mirrored pairs two signs occur more frequently in one location or occur only in one location of the Indus Valley, as shown in Table 3. The distances between these Indus Valley sites differ from 670 km to at least 900 km as seen in Fig. 4. We conclude that these signs are not mirrored signs, rather the directionality of these signs are due to infrequent usage and/or location.

Table 3 The location and frequency of two of the claimed mirrored signs.			
Sign	Locations		
	Harappa	Lothal	Mohenjo-daro
↕	1	-	6
↕	-	1	-
↕	11	-	-
↕	-	-	1

Sign. The mirrored sign occurs only once in Lothal, which is around 700 km away from the other locations which use the original sign . In addition, the pair of / occurs in Harappa which is similar to the one in Lothal. Therefore, we conclude that the mirrored sign found in Lothal is a location anomaly.

Sign. The mirrored sign occurs once in Mohenjo-Daro which is roughly 685 km from Harappa where the original sign occurs 11 times. We conclude that the reversed sign is a writing direction and a location anomaly. In addition, we state that the direction of the tick of sign 851 indicates writing direction.

Sign. According to ICIT, the mirrored sign occurs 10 times. In all of these 10 cases, we see that the mirrored sign is a terminal sign on the left side and that the next sign is always a form of . In the majority of the artifacts with the original sign, we see that the original sign is also a terminal sign and is followed by a variation of ; therefore, we conclude that the mirrored sign is negligible. Wells also noted that these two signs created bonded clusters (Wells, 2015).

A few of the artifacts may not be the mirrored sign or a unique variation. Tablet H-344 is unrecognizable, therefore, it may not contain the sign . Seal L-9 is broken, so the mirrored sign may not be on the artifact. Tablet H-2084 has the mirrored sign rotated 45°, which is a writing anomaly.

Sign. Two occurrences exist for the mirrored sign , one in Harappa and one in Mohenjo-daro. Out of all the seals with sign regardless of being the original or mirrored, Seal M-784 is the

Fig. 4 The red boxes highlight some of the major archeological sites of the Indus Valley Script inscriptions. The red boxes were added to a map from the Wikipedia article on “Indus Valley Civilization” and appears at <https://bit.ly/39B95az>. The original map was created by Avantiputra7 and is used by a CC BY-SA 3.0 license.

Table 4 The CISI tablets which are most similar to H-298, on the front and back.

CISI of artifact	Front	Back
H-298		
H-894		
H-1178		

most legible. The artifacts show that the sign has zero common signs other than ; therefore, the evidence shows that the direction of writing is negligible. Additionally, we see that the original and mirrored sign have an almost equal distribution.

Sign . The mirrored sign occurs only twice in comparison to the original which occurs 57 times. Rectangular Tablet H-298 is found in Harappa and reads . The original sign occurs frequently with , e.g. occurs on rectangular Tablets H-894, H-1178, and H-1946 which are all found in Harappa. Tablets H-894 and H-1178 are most similar to H-298 on both sides of the tablet as seen in Table 4. Due to the similarity on both sides of the seals and is found in the same location, the evidence shows that Tablet H-298 is an incorrect marking.

For the triangular Tablet H-1920 which reads the evidence shows that is mistakenly carved due to being rarely used in combination with signs and the use of the triangular tablet. The sign is often found as a terminal character on the left-hand side, which provides further evidence to neglect meaning behind the mirroring of the sign.

Sign . The original sign and the mirrored sign occur infrequently and all occurrences are located in Mohenjo-daro. There is an equal distribution of using the original or mirrored sign. In addition to this, we see that none of the seals contain any signs in common, regardless of containing the original or mirrored sign. Due to this, we state that this sign is a writing anomaly; therefore, directionality is negligible.

Sign . There are three occurrences of the mirrored sign which all occur as seals in Mohenjo-daro. There is no overlap between any of the seals that contain the original sign or the mirrored sign . The probability of using an original or mirrored sign has an almost equal distribution. We conclude that the directionality of the sign is negligible.

Reducing the Indus Valley Script sign set by data mining inspections

Reducing the Indus Valley Script sign set is important because it is the first step in the decipherment of the script. A sufficiently small set of signs is needed before the tedious guesswork of assigning phonetic values to each sign. If several variations of a sign are not recognized as the same sign, then it could lead would-be-decipherers astray as they assign different phonetic values to the different versions. Intuitively, a good grouping of signs would contain few singletons, that is, signs that occur only once in the corpus of inscriptions.

Many scholars have grouped similar signs together (Fairservis, 1983, 1992; Mahadevan, 1977; Parpola, 1994) but their groupings left a large number of singletons. Mahadevan claims to have found 417 signs. Out of these 417 signs, nearly 27 percent occur only once. Wells found 694 signs with 222 singletons. The correct number of signs may well be any number between 417 and 694, that is, it is just as likely that Mahadevan was too eager to group some signs together as Wells was too cautious and left separate

variations that really denote the same sign. In our grouping work, we started from the more expansive list of Wells because Mahadevan did not provide written reasons for his groupings. In contrast, we wanted to support our groupings with an explanation that can be checked and verified by future researchers.

After some experimentation, we identified six different reasons for grouping signs together. Identifying these reasons based on some early examples was helpful in debating whether some sign variations in later examples should be grouped together. We could go through our list of possible reasons and check whether each reason applied or not. This method sped up the decision process regarding grouping. We also believe that the method of stating explicitly a set of agreeable reasons for grouping makes the results more robust. That is because once people agree that something is a good reason for merging, the identification whether the reason applies or does not apply is often straightforward. Therefore, once the agreement on reasons is reached, people tend to come to the same conclusion regarding whether two signs should be merged or not. This methodical approach is better than having no explicitly stated reasons for merging but relying only on an implicit “visual sense” that may be subjective and different for each scholar. On the other hand, we admit that we do not present a fool-proof method for grouping signs, and future researchers may well identify additional reasons for grouping signs beyond the six reasons that we present below.

Six reasons for grouping signs. We found the following six reasons for grouping together signs:

Reasons for grouping signs

1. The signs are squished or morphed due to space issues.
2. The signs are mirrored without an underlying meaning.
3. The sequence that the main sign is found on is the same as the variation.
4. The sign is a location anomaly.
5. An incorrect sign is noted.
6. Visual similarity of signs is high and a varied sign occurs only once.

Below we illustrate these types of reasons in separate subsections. Table 5 shows the grouping of signs with 27 variations that could be merged with another sign, called the main sign, according to the six types of reasons.

Reason 1: Space issues lead to signs being squished. Signs which are squished or morphed due to space issues should be placed in the same set as the main sign. In this section, we describe, illustrate, and show proof for signs which fall into this type. The signs and are the same. There are three occurrences of ; however, upon further inspection, it is clear that the sign is only altered for space issues as seen in Fig. 5. Similarly, and are the same signs, but the second is more constricted. Signs and are also the same, as can be seen in H-1936, sign curves towards the bottom of the round seal.

Sign and should be the same sign. Figure 6 shows on the left side a normal sign, while on the right side the sign is missing the line in the middle due to the limitations of space. In addition to this, we see that the signs are all on the right.

Sign is a squished version of . Sign appears only once on a crowded seal and the evidence shows that it is the same as . In addition, we see that sign is incorrectly noted as a diamond pattern in the ovular shape when it is shown to be ovular inside as well, much like . Signs and are condensed versions of and . The seals where and appear are limited in space.

Table 5 The signs grouped according to the six types of reasons.		
Reason	Main Sign	Sign Variations
1		
2		
3		
4		
5		
6		

Reason 2: Signs are mirrored. Sign is uniquely found on a mirrored seal. As seen in seal H-2085, there are two mirrored signs found on the seal. The is mirrored on one side and is mirrored on the obverse. This would be according to the classification of Section 3 a Type 1 case for considering the and signs to be the same. Similarly, the mirrored signs of Section 3, except for the Type 5 case of and , can be also grouped together.

Reason 3: The sequence that the main sign is found on is the same as the variation. The signs grouped according to Reason 3 are shown in Table 6. It is possible to imagine that these pairs of signs serve some function such as different plural markers for male,

Fig. 5 Constriction of the man-with-comb sign. Seal 241 from Vats (1940) shows that the sign is constricted.

Table 6 The signs found in similar position patterns, where the "X" can be replaced with any of the signs shown in the first column.	
Signs	Sequences
and	X
	X
and	X
	X
and	X
	X
, and	X

female and neutral genders. However, scripts that use different grammatical markers, such as Minoan Linear A and Mycenaean Greek Linear B, usually denote those markers by completely different signs instead of by signs that have only tiny differences. Moreover, the variations occur at various places (initially, medially, and finally) within these short sequences, which are presumably single words. Hence if these signs denoted grammatical markers, then they would be unlikely to all just denote one type of marker, such as gender. We would need to consider several types of grammatical markers that all happen to be denoted by tiny differences in signs. That seems improbable.

Reason 4: The sign is a location anomaly. Signs and are noted to be the same sign. Sign is found in Lothal which is nearly 700 km away from Mohenjo-daro and Harappa where the sign is found.

Fig. 6 Seals 99 (left) and 68 (right) from Mackay (1938). The sign shown on the left seems to be simplified to on the right because of space limitation.

Fig. 7 Incorrectly carved bow sign. Seal 230 from Vats (1940) shows that the supposed sign is a sign that only slightly extends the bounds.

Reason 5: Incorrect sign noted. Human error was possible when these signs were carved or when they were inserted into the ICIT database (Wells and Fuls, 2017). We see that signs , , and are the same sign, and are incorrectly noted in the dataset. In comparison we see that in sign the curve has slightly gone out of the bounds on the left bottom side as shown in Fig. 7. Therefore, the evidence shows that this is a carving mistake.

In addition, we see that sign is incorrectly noted as a different sign from as the signs were the same according to CISI on seal M-109; however, we see that seal C-9 appears skewed to the top where the sign is cut off. Similarly, we see that sign is only slightly shifted on the left side, which is due to carving error. Sign occurs on two items, a bangle where the seals are curved due to the nature of the object, and on a seal which cannot be validated due to only a drawing existing for it in CISI; therefore, we conclude it is an error in notation. Finally, the evidence shows that sign has been incorrectly noted, and the seal that supposedly contains this sign actually contains three signs, which includes instead of .

Reason 6: Visual similarity of signs is high and varied sign occurs only once. In addition to the grouping of signs which we have shown evidence for, we speculate that the following few signs are the same due to the extreme similarity of the signs and due to that the variation occurs only once.

Sign occurs once in comparison to sign which occurs eight times. The only difference between the two is the line above the diagonal. We see that sign occurs in Kalibangan on a pot, which is evidence for being an anomaly in terms of location.

Signs , , and all occur only once. They are all found in Mohenjo-daro and have not occurred in the other locations of the Indus civilization. Due to being not only a location anomaly but an outlier in Mohenjo-daro, the evidence shows that the signs should be considered the same.

Sign occurs once in comparison to sign which occurs nine times. The sign appears constricted on the top of the seal and in the corner. Due to having such high similarity with , the sign seems an erroneous version of sign .

Conclusions and future work

Even though there have been over 100 decipherment attempts of the Indus Valley Script, none of the attempts involved a thorough examination of redundancies in the sign lists. This article identified 50 signs as redundant in the Indus Valley Script sign list of Wells (2015). Hence would-be-decipherers need to find the correct phonetic values for 50 fewer signs than they previously thought. That is a significant reduction in the search space for phonetic assignments to the Indus Valley Script signs.

This methodology could be expanded to other anomalies found in the Indus Valley Script and lead to a greater understanding of this undeciphered script. Analyzing the location, the context (Ansumali Mukhopadhyay, 2019; Rao et al., 2009a, b; Yadav et al., 2010), the period of time when a sign was used (Possehl, 1996), or the object a sign was found on and the relationship of the Indus Valley Script signs to the signs in related scripts (Daggumati and Revesz, 2018, 2019; Hunter, 1929; Kak, 1988; Revesz, 2016a, b, c, 2017a, b) could facilitate a decipherment of the Indus Valley Script (Robinson, 2002, 2015).

Data availability

The Indus Valley Script image dataset analyzed during the current study is available in the original three volumes of the *Corpus of Indus Seals and Inscriptions* (CISI) (Joshi and Parpola, 1987; Parpola et al., 2010; Shah and Parpola, 1991), which contain copyrighted images. To illustrate some of the key concepts, in the figures we used some images of the Indus Valley Script seals and

tablets from Mackay (1938) and Vats (1940), which have copy-right free images.

Received: 17 February 2020; Accepted: 11 January 2021;

Published online: 17 February 2021

References

- Ansumali Mukhopadhyay B (2019) Interrogating Indus inscriptions to unravel their mechanisms of meaning conveyance. *Palgrave Commun* 5:73
- Daggumati S, Revesz PZ (2018) Data mining ancient script image data using convolutional neural networks. In: Flesca S, Zumpano E, Masciari E, Caroprese L (eds) Proceedings of the 22nd international database engineering & applications symposium. ACM, pp 267–272
- Daggumati S, Revesz PZ (2019) Data mining ancient scripts to investigate their relationships and origins. In: Desai BC, Anagnostopoulos D, Manolopoulos Y, Nikolaidou M (eds) Proceedings of the 23rd international database engineering & applications symposium. ACM, pp 209–218.
- Daniels PT, Bright W (1996) The world's writing systems. Oxford University Press
- Fairservis WA (1983) The script of the Indus Valley civilization. *Sci Am* 248(3):58–67
- Fairservis WA (1992) The Harappan civilization and its writing: a model for the decipherment of the Indus Script. Brill.
- Fuls A (2013) Positional analysis of indus signs. *Epigrafika* 7:253–275
- Hunter GR (1929) The script of Harappa and Mohenjodaro and its connection with other scripts. Ph.D. thesis, University of Oxford.
- Joshi JP, Parpola A (1987) Corpus of Indus Seals and Inscriptions. 1. Collections in India. In: *Annales Academiae Scientiarum Fennicae, Series B, Vol. 239*. Suomalainen Tiedakatemia, pp. 1–375.
- Kak SC (1988) A frequency analysis of the Indus script. *Cryptologia* 12(3):129–143
- Mackay EJH (1938) Further excavations at Mohenjodaro. Munshiram Manoharlal Publishers Pvt. Ltd., Delhi
- Mahadevan I (1977) The Indus Script: texts, concordance and tables, memoirs—Archaeological Survey of India, no. 77. Archaeological Survey of India
- Parpola A (1986) The Indus script: a challenging puzzle. *World Archaeol* 17(3):399–419
- Parpola A (1994) Deciphering the Indus script. Cambridge University Press
- Parpola A (1996) The world's writing systems. Oxford University Press
- Parpola A, Pande BM, Koskikallio P (2010) Corpus of Indus Seals and Inscriptions, vol. 3. New material, untraced objects, and collections outside India and Pakistan. Suomalainen Tiedakatemia
- Possehl GL (1996) Indus Age: the writing system, vol. 3. University of Pennsylvania Press
- Rao RP, Yadav N, Vahia MN, Joglekar H, Adhikari R, Mahadevan I (2009a) Entropic evidence for linguistic structure in the Indus script. *Science* 324(5931):1165–1165
- Rao RP, Yadav N, Vahia MN, Joglekar H, Adhikari R, Mahadevan I (2009b) A Markov model of the Indus script. *Proc Natl Acad Sci USA* 106(33):13685–13690
- Revesz PZ (2016a) Bioinformatics evolutionary tree algorithms reveal the history of the Cretan Script Family. *Int J Appl Math Inform* 10:67–76
- Revesz PZ (2016b) A computer-aided translation of the Cretan Hieroglyph script. *Int J Signal Process* 1:127–133
- Revesz PZ (2016c) A computer-aided translation of the Phaistos Disk. *Int J Comput* 10:94–100
- Revesz PZ (2017a) Establishing the West-Ugric language family with Minoan, Hattic and Hungarian by a decipherment of Linear A. *WSEAS Trans Inf Sci Appl* 14:306–335
- Revesz PZ (2017b) A translation of the Arkalochori Axe and the Malia Altar Stone. *WSEAS Trans Inf Sci Appl* 14(1):124–133
- Robinson A (2002) Lost languages: the enigma of the world's undeciphered scripts. McGraw-Hill, New York
- Robinson A (2015) Ancient civilization: cracking the indus script. *Nat News* 526:1
- Shah SGM, Parpola A, Stacul G, Lyytikä J, Ilyas SM (1991) Corpus of Indus Seals and Inscriptions. 2. Collections in Pakistan. In: *Annales Academiae Scientiarum Fennicae, Series B, Vol. 240*. Suomalainen Tiedakatemia, pp. 1–448
- Vats MS (1940) Excavations at Harappa. Munshiram Manoharlal Publishers Pvt. Ltd., Delhi
- Wells B (1998) An introduction to Indus writing. University of Calgary
- Wells B, Fuls A (2017) Online Indus writing database. <http://www.indus.epigraphica.de/>
- Wells BK (2011) Epigraphic approaches to Indus Writing. Oxbow Books
- Wells BK (2015) The archaeology and epigraphy of Indus writing. Archaeopress
- Yadav N, Joglekar H, Rao RP et al. (2010) Statistical analysis of the Indus script using n-grams. *PLoS ONE* 5(3):e9506

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to P.Z.R.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021