The Formal Economy of Written Signs

E. J. W. Barber

Those portions of a sign's shape which are distinctive, yet in themselves valueless, can be viewed as forming a system. Such systems have ranged historically from the very loose to the very tight. When the possibilities of arranging these elements are also considered, the great range in economy of design becomes even more apparent. Our own roman script, in fact, is rather inefficient in these terms; Morse Code, on the other hand, is quite efficient, though not perfect. It might prove useful to develop other scripts based on these principles of internal economy. But economy of form is not the same as efficiency of use: it is merely one possible component of efficiency, and must be carefully distinguished as such.

Much has been made of the relative economy of alphabets over syllabaries, logographic scripts, etc., for representing a language: alphabets require the fewest different symbols. But little has been written about the economy of purely formal design within a script—about the use of distinctive design elements and of patterns of arranging them. One might assume that the fewer the signs, the greater the economy of formal design would be, but historically that has not always been the case. Nor should economy of design be confused with problems of efficiency. Efficiency implies a purpose; and what is efficient for one purpose (e.g., writing) may be very inefficient for another purpose (e.g., reading). Economy of formal design, on the other hand, is an entirely self-contained measure, concerning merely the number of basic elements and patterns present. In order to see what is involved in formal economy, let us consider the structural principles of symbols and then some historical examples.[1]

Each sign in a script can be viewed as composed of elements which are distinctive—that is, serve to distinguish one kind of sign from every other—yet are in themselves without any direct representational value. These basic elements occur in various arrangements, each unit thus formed corresponding to some value.[2] Consider, for

Visible Language, VII 2 (Spring 1973), 155–166.
© 1973, *Visible Language*, c/o The Cleveland Museum of Art, Cleveland, Ohio USA 44106.
Author's address: Department of Linguistics, Occidental College, Los Angeles, Cal. 90041.

example, the Cyrillic letters ш (*sh*), щ (*shch*), ц (*ts*), and п (*p*). They are composed entirely of short and long vertical and horizontal lines—the presence, number, and position of which distinguish the letters from each other. But none of these single composing lines has a phonetic value of its own which it carries with it.[3] In short, the letters are the minimal units with linguistic values, while there are smaller critical units with no such values. The economy of the formal design of a script can thus be measured in terms of the number of these minimal critical units used, and also in terms of the number and complexity of patterns in which they are arranged.

The history of writing systems shows that most ancient scripts developed from pictorial sign forms of some sort, rather than from abstract marks.[4] Since the scripts did not originate as abstract contrastive systems (although pictorial signs were generally reduced to considerable simplicity), there is no particular reason to expect great economy of *system* in the critical elements and their arrangements. In many cases no such economy ever developed; but in others, particularly where a medium less versatile than pen and ink prevailed, varying degrees of economy of form in fact arose.

Perhaps the earliest script to attain a considerable measure of formal economy was cuneiform. The Sumerians began, like most people, with pictorial signs representing such common sights as fish and grain (Fig. 1): complex shapes in comparison with the abstract world of circles and crosshatches. But the chief medium of writing in mud-bedded Mesopotamia was clay. Unfortunately, if one line is required to cross another, scratching the second line in this medium will cause partial or complete obliteration of the first, for little rolls of clay shoved aside by the scratching point will be pushed into the furrow which forms the first line. So the Sumerians eventually resorted to the technique of pressing rather than scratching the lines. The shapes were thus gradually converted from recognizable curvilinear pictures to groups of short, straight, impressed lines (cf. Fig. 1). Because of the shape of the stylus used to form them, moreover, these lines came to have more the appearance of little wedges than of little lines (hence the term "cuneiform," or "wedge-shaped"). By the late third millennium, the signs had become pure patterns of wedges.

These wedges constitute an extremely economical system of minimal critical writing units—units corresponding to no linguistic value

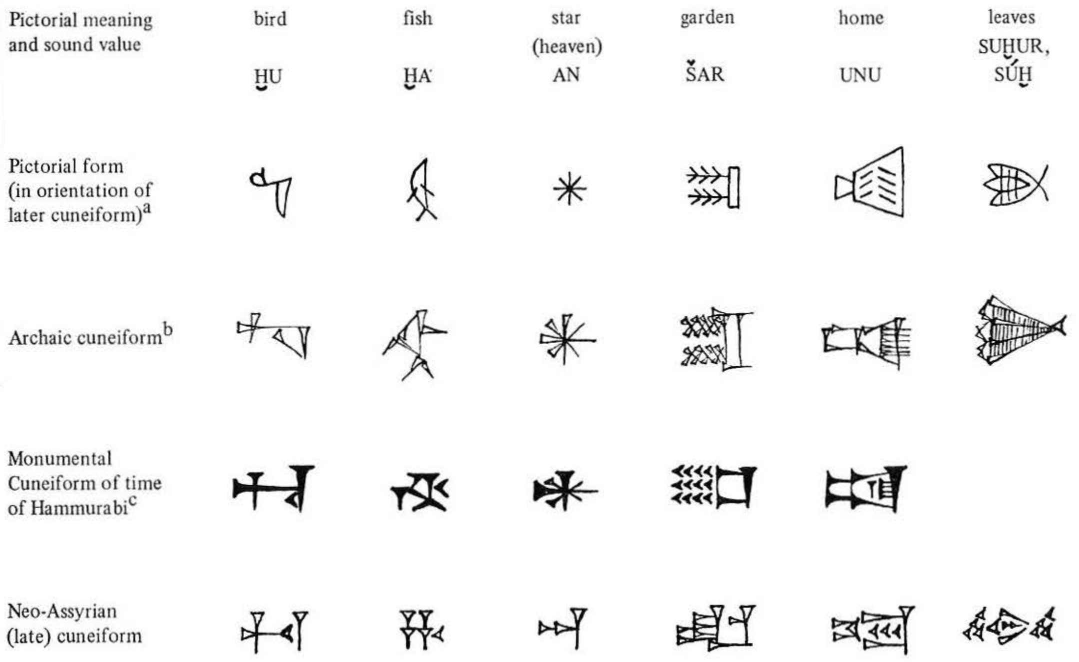

Figure 1. Development of Cuneiform.
a. From Uruk and Jemdet Nasr; as shown by A. Falkenstein, *Archaishe Texte aus Uruk*; Berlin, 1936.
b. From Uruk and Fara; as shown by Falkenstein.
c. As shown by R. F. Harper, *The Code of Hammurabi*; Chicago, 1904.

and yet serving to distinguish one sign from another. From Old Babylonian to Neo-Assyrian times, the signs consist exclusively of two types of wedges, tailed and tailless, of a small number of subvarieties. Tailed wedges occur with tails straight down, slanting down to the right, horizontal to the right, and (more rarely) slanting up to the right:

Wedges pointing straight up or to the left have dropped out: evidently it was a nuisance to have to turn the stylus that far. In some periods and areas the tails even seem to occur distinctively long and short, or large and small. Tailless wedges, too, occur large and small.

(Examples of almost all of these types can be seen in Fig. 1.) Thus an extremely small number of elements comprise the basic building blocks from which hundreds of signs are composed.

On the other hand, when we turn to consider the patterns according to which these few wedges are arranged, the extreme economy breaks down. Nothing remotely approaching the total number of patterns simpler than the most complex patterns in use can be found in the script. Numerous simple patterns are missing, whereas other signs are wildly complicated. For example, the combination [a] is not found in Old Babylonian, although both [b] and [c] occur (as the numeral "1" and the syllable "tab," respectively), and the combination occurs in, among others, the sign [d]. Of course, the

 [a]  [b]  [c]  [d]

historical reason for this looseness of form is obvious: the signs grew up more or less accidentally from pictures, instead of being invented from the start as abstract patterns of wedges.

Of all the prealphabetic scripts that I know of, cuneiform has the greatest claim to economy of form, even though this economy is chiefly in the distinctive elements rather than in the patterns. Egyptian, Hittite, and Mayan hieroglyphics—each having, like cuneiform, several hundred signs apiece—are composed of countless types of lines, squiggles, and curliques. Considerations neither of speed nor of medium[5] ever forced them to abandon such wealth of form. The Chinese, on the other hand, evidently found the paintbrush a not entirely versatile tool, and reduced the composition of their thousands of characters to a handful of brush strokes—more types of strokes than there are wedges in cuneiform, but still significantly few. And as in cuneiform, there is no tight economy of pattern.

With the advent of purely phonological scripts, in which every sign represented a sound or sound combination rather than a word, the number of signs was reduced from hundreds and even thousands of signs to well under one hundred (the exact number varies with the script). The chances of developing a script economical in terms of pattern simultaneously increased; but such a development, even then, did not often occur.

Figure 2. Ugaritic Cuneiform. After "Textes en cunéiformes alphabétiques," *Palais Royal d'Ugarit* II (ed. C. F. A. Schaeffer); Paris, 1957; pp. 199–202.

The earliest phonological script of noteworthy economy is the cuneiform "alphabet" used in the late second millennium B.C. at the ancient city of Ugarit on the Syrian coast (Fig. 2). Although cuneiform in method of production, this group of thirty signs is not historically derived from the complex Mesopotamian script. The signs were evidently invented afresh, perhaps even to be simple abstract patterns. But although the maximum number of wedges per sign is

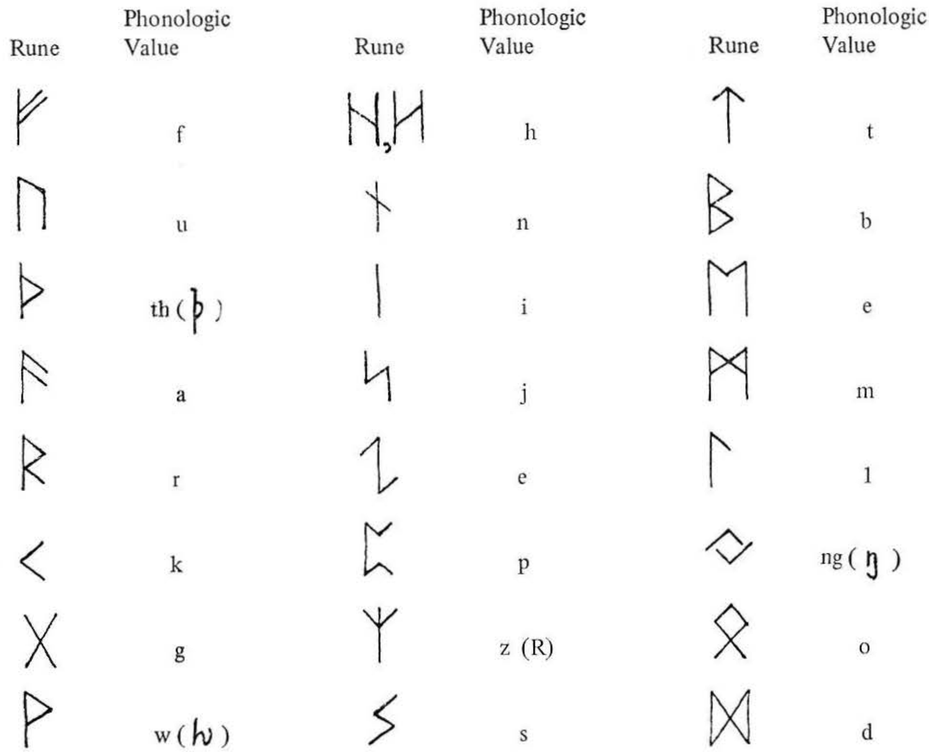

Figure 3. Early Runic Alphabet (in traditional order).

cut down to a mere seven, still not all the patterns simpler than the most complex are used. There is no sign consisting of four verticals, for example, although there is one with six verticals.

Another notably simple ancient system is the early runic alphabet, used in the mid-first millennium A.D. in Scandinavia (Fig. 3). The signs are composed entirely of three sorts of straight lines, each occurring both long and short: lines slanted to the right and to the left, and verticals. Once again this simplicity of design is the result of the historical medium, it seems, for the runes were originally slashed into wood with a straight knife.[6] The complete lack of horizontal straight lines is usually explained by the suggestion that if the grain of the dressed wooden plank is running horizontally, a horizontal slash would be hard to distinguish from the grooves in the grain of the wood—in addition to the fact that a knife blade would get caught

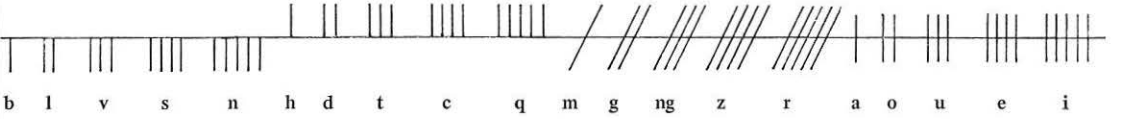

Figure 4. Ogham Alphabet.

in the grain and be forced out of the desired path. As long as the knife slashes across the grain, either directly or at a considerable angle, the grain presents no problem. But despite the economy of the component lines, the runic alphabet still is not particularly economical when it comes to patterns. Instead it was the Celts, living to the west of the Germanic rune-writers, who came up with the ultimate in economy of form in ancient scripts and did so with letter forms intended, like runes, to be slashed.

The Ogham script (Fig. 4) is, all told, perhaps the most economical ever used before modern times. Each sign consists of one to five straight and identical lines, the lines being of four sorts: (1) to one side of the median, (2) to the other side of the median, (3) crossing the median—all three of these types being short and at right angles to the median—and (4) crossing the median but long and slanted. (The "median" was generally the running edge of a squared timber or block of stone, the signs bending from one face to the other.) Since all combinations are used, these few patterns and line types account exactly for the twenty signs of the script. The only superfluous feature is in the last category of signs, since either slanting or length technically would have been sufficient to differentiate this series; the inventor need not have added both.

Still more economical in terms of elements, though less so in terms of patterns, is modern international Morse Code. We usually think of it as visible dots and dashes, but the sign system is predicated solely on the opposition of long against short; it can be used in any medium where such an opposition can be defined—sound, light, electricity, or whatever.[7] It thus consists of one pair of oppositions (long/short), whereas Ogham had consisted of a three-way opposition (one side/other side/crossing) and two pairs of oppositions (long/short, perpendicular/slanting).[8] When we come to consider patterns, however, standard international Morse Code falls slightly short of true economy. All of the simple patterns are used: all combinations and distinguishable permutations of dots and dashes taken up to four at a time are used (Fig. 5), with the single exception of four dashes. But since there are only thirty such combinations, and Morse Code accommodates 52 signs (including punctuation and digits), obviously the code has to spill over into the next series, namely taking dots and dashes five at a time. Instead of doing that, it uses only fourteen of these 32 possibilities, and for the remaining nine uses combinations of six dots and dashes, a considerable breach of economy.

The most economical of all in terms of elements is Braille, which uses only dots. These occur in combinations of one to six, permuted in six positions. In addition all 63 possible combinations are used, one way or another,[9] although a few are used purely as signals to alter the value of the succeeding block of dots. The use of these signals vastly increases the flexibility of Braille, but decreases its overall economy, since relatively few two-block groups have special values.

By the criterion of economy which we have been developing here, our roman script is a disaster. Like most scripts, it developed from pictorial signs, and the fact that it remained chiefly a pen and ink script through all but its most recent history long shielded it from having to become more economical in design. For a while in the Middle Ages, indeed, the roman script became less and less economical in form, as anyone who has tried to read the average lot of old manuscripts will ruefully testify. It was chiefly the invention of printing that reversed the trend and brought the script back nearly to the simplicity of the classical Roman source. But even its current form has not been sufficiently economical to accommodate either Samuel Morse and his telegraph or the modern computer engineer. Both were driven to the expedient of inventing entirely new systems, which recode the alphabet altogether. Recent attempts, on the other hand, to redesign the roman script to make it directly readable by computers have generally succeeded in reducing the number of minimal critical elements to a handful (Eden, for example, reduces cursive to four).[10] Perhaps economy of formal design is the direction in which we are now heading.

On the other hand, I must emphasize once again that economy in the sense that we have been discussing is not to be confused with efficiency. If a script is being designed to be practical, there are

1:	•	E	–	T
2:	••	I	––	M
	•–	A	–•	N
3:	•••	S	–––	O
	•––	W	–••	D
	••–	U	––•	G
	•–•	R	–•–	K
4:	••••	H	(––––	does not occur)
	•–––	J	–•••	B
	••––	Ü	––••	Z
	•••–	V	–––•	Ö
	•–••	L	–•––	Y
	•––•	P	–••–	X Q
	••–•	F	––•–	Q
	•–•–	Ä	–•–•	C
5:	•••••	5	–––––	zero
	•––––	1	–••••	6
	••–––	2	––•••	7
	•••––	3	–––••	8
	••••–	4	––––•	9
	••–••	È	––•––	Ñ
	•––•–	Á	–••–•	/

(remaining 18 patterns of 5 dots and dashes do not occur)

6:	•––––•	’	–••••–	hyphen
	••––••	?	––••––	,
	•–•–•–	period	–•–•–•	;
	••––•–	dash		
			–•––•–	parenthesis
			–––•••	:

(remaining 55 patterns of 6 dots and dashes do not occur)

Figure 5. The 52 Signs of International Morse Code Arranged According to Internal Pattern.

several important considerations beyond mere economy of form—although an analysis of the form often helps in understanding the broader concept of efficiency. For example, in Morse Code a dash takes so much longer to transmit than a dot that it is more efficient in some media, especially where time and electricity are concerned, to ignore the combination of four dashes in favor of five dots.

Still more important than economy of materials, in most cases, are problems of human perception. Precisely because Morse Code is so economical in form, it lacks the redundancy which we humans use to catch our own mistakes. Thus Morse Code can be difficult to perceive accurately in a hurry or in the face of outside interference. And if the human decoder should make a mistake in perception—say, picking up three dots where there were really four, or inverting the order of a dot and dash somewhere—the result is a totally unrelated member of the alphabet: S (···) for H (····), or B (-···) for L (·-··). Sometimes the high redundancy of human language can be used to straighten things out again ("Send a salf ton of coal"); but where it does not, the code gives no clue that anything is even wrong, let alone how to fix it ("Send three babes to the warehouse"). There is thus great advantage to be gained, when human fallibility is concerned, by using a system which is not so economical. If the form "ɵ" turns up, in our Roman script, we instantly know something is wrong, because no such combination occurs, although circles and straight, horizontal lines occur. Furthermore, we can narrow our search for a remedy down to two most likely letters: "o," on the hypothesis that an extra horizontal crept in, and "e," on the hypothesis that the circle somehow got closed. A certain flouting of economy thus has its place, both in speed of perception and in trouble-shooting.

Overall efficiency and economy can be affected by yet another design consideration. Just as big words are sometimes built out of little words, such that the meaning of the big word is at least partly composed of the meanings of the little words (e.g., "spoonful"), so complex meaningful signs can be formed out of minimal meaningful signs. The Ogham script, alone of all the ancient scripts that I know of, took this principle into account, by putting all the vowels into one series of signs. Thus the series of one to five short lines crossing the median at right angles represents all the vowels and only the vowels. One could therefore reanalyse the script to say that being short and crossing perpendicularly indicated a vowel, and that the number of lines indicated which vowel. The Celts did not carry this composition principle to the extreme—if the consonants are organized, the basis is not clear—but we could carry it out.

Suppose, for example, that one reorganized and elaborated upon the economical Ogham system to produce a phonetic (or, for that matter, a phonemic) notation. It would not be efficient for hasty field transcription, but it could be quite revealing as an analytical tool under less hectic conditions. Voiced sounds might occur on one side of the median, voiceless sounds on the other: the analyst could then see at a glance the structure of the language in terms of voicing. Suppose that stops were represented by long lines, spirants by short ones, and resonants by dots: the syllable and cluster structures of the language would again be evident at a glance. The number of lines might code the distance of the articulation from one end of the vocal tract (e.g., one for lip articulation, two for dental, etc.). With the system worked out in full, a particular sound could be recorded by a particular "sign," where the sign was made up of lines or dots. But each pattern, type, and number of the formal elements within that sign would carry a component meaning. Such a script would indeed have a kind of ultimate economy of form, and for its limited purpose (leisurely phonologic analysis) even be quite efficient.[11]

It would be interesting to develop other scripts to be as economical as possible for the contexts in which they are to be used. A code to be used entirely within a machine, for example, or one which is not intended for rapid human use, can afford to be less redundant than one to be used in daily life. Current efforts to develop "self-explanatory" codes to send out into space, in hope of reaching other intelligent beings, represent yet another interesting branch of this same problem. Unfortunately, many of the factors which have been cited here as contributing to ultimate efficiency, as opposed to mere economy of form, are difficult to measure or are virtually unexplored. The problems of human perceptual limitations, moreover, are perhaps the least known right now. Nonetheless, it might be worth the time and trouble to investigate thoroughly the economy and efficiency of the roman script—by mere accident of history the chief vehicle of communication in the most hurried society the world has known.

1. Elaborated from material in the author's forthcoming book, *Archaeological Decipherment*, to be published by Princeton University Press in 1973.
2. Such an analysis will be familiar to linguists, as being parallel to the distinction between phonemes (minimal distinctive units, but without meaning) and morphemes (minimal units with meaning).
3. In the Ukrainian version of Cyrillic, it is true that a single vertical stroke has the value *i*, all by itself; but it does not maintain that value within the above-mentioned letters. Thus ш still represents *sh*, not *iii*.
4. Cf. I. J. Gelb, *A Study of Writing*, Chicago, 2nd ed. 1963.
5. Quite probably the chief medium for each was pen and ink, although that has not been proved in every case.
6. When occurring in other media, some of the letters also show more curved forms. But the angular forms seems to predominate.
7. It is interesting that, although other writing systems too use the contrastive opposition of sign vs. space (or blank), Morse Code is so economical in its system of oppositions that the entire system, even including the spaces, reduces to only two oppositions. The underlying system is merely "something" (sound, light, electricity, or whatever) vs. "nothing" (pause, blank, etc.) ,each occurring short vs. long, measured appropriately in time or space.
8. Since, as in Morse Code, the opposition short/long is also used in the spacing of lines within and between signs (short within and long between), dropping the opposition of perpendicular/slanting would do most to tighten the formal economy of Ogham.
9. At least, as commonly used in the U.S. Some represent the 26 letters of our alphabet, others punctuation and diacritics, still others frequent combinations of letters (ligatures, as it were). The few left over—purposely left, it seems, because they are too difficult to tell apart from their mirror-image counterparts—are used as signals to alter the value of the succeeding sign. Counting these signals, every combination is used.
10. Murray Eden, "On the Formalization of Handwriting," *Structure of Language and its Mathematical Aspects* (*Proceedings of Symposia in Applied Mathematics* XII); American Mathematical Society, Providence, 1961; 83–88. Although Eden reduces the letters to forms containing only these four distinctive strokes, the forms are not necessarily instantly recognizable to one who knows only ordinary American cursive script. (In the same way a syllable-initial American *r*, stripped of its characteristic, but nonphonemic, lip-rounding is generally unrecognizable in initial position to the average American.) This wealth of characteristic but nondistinctive forms adds enormously to the lack of economy in roman script.
11. Henry Sweet, father of modern phonetics, attempted a structural notation system much along these lines, although with not quite the degree of economy suggested here. See Sweet, *Primer of Phonetics*; Oxford, 1890.