Review of *Numerical Notation: A Comparative History,* by Stephen Chrisomalis, Cambridge University Press, 2010. ix+486 pps.

After the development of words for numbers, the construction of a written notation for numbers is one of the most fundamental and most widespread initial steps in the development of mathematics. Numerical notations can be found on artifacts from both Egypt and Mesopotamia from about 3200 BC; in Egypt concurrent with the earliest written records of language, and in Mesopotamia predating them by centuries. But it is possible that numerical notations are in fact much older than that: there are marks on Neolithic pottery and tortoise shells found at Jiahu in Henan province dating from 6600-6200 BC that may well be numeral signs.

Among cultural artifacts, numerical notations are unusual in two respects. First, their semantics is extremely well-defined and limited; mostly, they represent some or all of the natural numbers, sometimes with some fractions. Second, they are not very numerous; there are records of only about 100 distinct systems that have ever been in any kind of general use. (The exact count depends on how you individuate.) It is therefore possible to present an account of numerical notational systems throughout history that is essentially complete and definitive, up to the limits in the historical record.

Stephen Chrisomalis' *Numerical Notations: A Comparative History* is such an account. He gives a complete description of all the numerical notation systems that are known: how they worked and how they were used. Each system is illustrated with a clear hand-drawn table of symbols; most are also accompanied by a photograph of their use on a historical or archaeological artifact.[1] Chrisomalis also traces the evolutionary history of these systems; that is, how one notational system evolved out of another. When, as often, this history is obscure, he surveys the scholarly literature, and makes his own judgment of the probably historical relations between notations, taking into account such considerations as similarity of structure, similarity of symbols, known contacts between cultures, and proximity in time and space.[2] After completing the survey, he analyzes a substantial number of regularities and near regularities that govern individual notations, and a much smaller number of regularities that govern how one numerical system can evolve out of another. Finally, he discusses how these regularities relate to characteristics of human cognition and human society.

Chrisomalis classifies the structure of notational systems into five main categories. All numerical notations are built around powers of a fixed base, generally but not always 10. In a *cumulative-additive* system, such as Roman numerals, there is a symbol for each power of the base (I, X, C, M); these are repeated; and then the values are added. (The Roman numerals also have a *subbase* of 5 (V, L, D), common in additive systems, and a *subtractive* feature (IX for 9) which is almost unique.) In a *ciphered additive* system, such as the Greek or Hebrew numerals, each multiple of a power of ten has its own symbol, and the values of these are added together. For instance, in the Greek alphabetic system, ν represents 400, λ represents 30, and δ represent 4; thus $\nu\lambda\delta$ represents 434. In a *multiplicative-additive system* signs for the digits 1-9 alternates with signs for the power of 10. The traditional Chinese numerical notation works this way; so (to some extent) does the English language, e.g. "two *thousand* three *hundred* for*ty* seven". In a *ciphered positional* notation, such as the Western numerals, there are symbols for the numbers from either 0 or 1 up to the base minus 1; the power off the base is then indicated by the position of the symbol in the numeral. To be unambiguous (not all such systems were) there must either be a symbol for zero or some other way of indicating powers with a zero coefficient. Finally a *cumulative-positional* system, powers of ten are represented positionally, as in the Western numbers, but their coefficients are represented cumulatively; the famous base-60 ancient Babylonian system followed this principle. (A ciphered base-60 system would of course need 60 distinct symbols for the digits.)

[1] In the online copy that I was reading, these photographs were not always very clear; but the reproduction quality may well be better in printed copies.

[2] Some of the scholarly literature area has wildly conjectured relations between notational systems that are similar in some respects, but separated by millennia.

About 30 percent of the systems that Chrisomalis discusses are *hybrids* that combine different principles for different ranges of numbers; often, a cumulative or ciphered additive system for lower powers of the base and a multiplicative additive system for higher powers. However, no naturally-arising systems of pure numbers use any other principles. One can imagine a numerical notation that represents numbers by their prime factorization; or that uses division (as in "a half-dozen"); or that uses the factorials as a base (e.g. representing 301 as [2,2,2,0,1] since $301 = 2 \cdot 5! + 2 \cdot 4! + 2 \cdot 3! + 1 \cdot 1$) etc.; but these do not actually arise.

Chrisomalis' historical accounts are always impeccably clear, but unavoidably somewhat dry; after 100 numerical notational systems, one's eyes do begin to glaze. However, there are all kinds of fascinating historical and cultural tidbits to enjoy on the way. Large numbers and wild exaggeration using large numbers go back to the very earliest days of numerical notation; an Egyptian mace-head from 3100 BC records the supposed capture of 120,000 prisoners, 1,422,000 goats, and 400,000 cattle. Some numerical systems were used only for counts of quite specific categories; in fact in ancient Uruk in Sumeria, there were 15 different numerical systems for different kinds of quantities, including "the regular Š system [for] barley, the Š' system for germinated barley for brewing beer, and the Š* system for barley groats." In modern China, there are six numerical systems that are to some degree active, depending on the region and the particular use. One of these, the "accountants' system" uses deliberately complex symbols, in order to avoid falsification.

The historical understanding of the origins of the Western numerals, which originated in India and were transmitted via the Islamic world to Europe, deteriorated over the centuries. In the earliest European sources, as well as in Arabic sources, they are, correctly, called "Indian" numerals. By the sixteenth century, they were often called, less correctly, "Arabic" numerals. In the early twentieth century, a number of scholars proposed, on the basis of no evidence other than pure Eurocentrism, that they must actually have originated in Greece.

Chrisomalis emphasizes strongly that the use of numerical notations varies significantly from one culture and time to another, and that we should not make the mistake of supposing that our own uses of numbers apply universally. In particular, in most times and places, written numbers were not used for calculation; calculations were done using some method of finger calculation or using tools such as an abacus or counting sticks. It is therefore a mistake to suppose that the inefficiency of a numerical notation for doing calculation was any kind of a drawback.

By way of analogy (mine, not Chrisomalis'), consider the numerical notation for dates e.g. 2/24/2015 for today's date February 24, 2015 (American style). What is it good for? Well, it is easy to approximate the time gap between dates, particularly if they don't cross a boundary: 7/15/2015 will be about 5 months from now; 9/26/1898 was about 117 years ago. It is also easy to judge the relation of dates to yearly events such as holidays, some astronomical events, and seasons: 12/25/1898 was Christmas, was about 4 days after the winter solstice, and was the beginning of winter in New York and the beginning of summer in Sydney. That's about it. Calculating what day of the week was 12/25/1898 or exactly how many days passed between 12/25/1898 and 2/24/2015 is laborious by hand, and requires a several line program to do by computer (and you have to be very careful to avoid off-by-one errors.) And the irregularity of the calendar is a constant source of inefficiency and trouble in making calendars, either printed or automated. Why do we put up with this? First, because we rarely have to compute the number of days since 12/25/1898 (though we do often have to determine the day of the week of some future date). Second, because the costs of changing it would be prohibitive. Third, because an earth year happens to be 365.2425 earth days, so no calendar that incorporates both years and days can possibly be very elegant.[3] If our descendants living on seasonless space stations make fun of us for using such an obviously awkward

[3]It seems to me, by the way, that this is a clear counter-example to the common theory that we find mathematical regularities in nature because we impose them as a conceptual framework. If we could impose preferred mathematical regularities on nature, we wouldn't be dealing with this.

way of measuring time, they will simply be missing the point.

Even in the basic Western numerical notation, there are suboptimalities that we tend to overlook because we are so used to them. God only knows how many man-hours and dollar-equivalents have been lost over the last five centuries due to the fact that in many handwritings the digits '4' and '9' and the digits '1' and '7' are easily confused. The Roman numerals are much clearer in that respect.

To me, the most interesting part of Chrisomalis' book is his analysis of the regularities that govern numerical systems. He adduces 14 principles that hold in all the systems he has studied and 8 that holds in nearly all. Examples of universals: "Every base is a multiple of 10"; "Any system that can represent $N + 1$ can also represent N." Examples of near-universals: "No numerical notation explicitly represents arithmetic operations such as addition and multiplication" (contrast linguistic forms such as "a thousand *and* fourteen" "vingt *et* un"). The single exception is in the Shang Chinese numerals. "All numerical notation systems are ordered and read from the highest to the lowest power of the base". This is, of course, necessarily true for positional systems, but would not have to be true for additive systems. One could imagine, in Roman numerals, that you could write IICVCX to mean 217; but in fact this is not allowed. There are a few exceptions in some alphabetic system, where the notation follows the word order for the lexical number. Chrisomalis' explanations of these in terms of human cognitive capacity, such as the limited size of working memory, and the relation of numerical notations to language, are thought-provoking and no doubt true in part. I don't think that they are a sufficient explanation of all the regularities found, but probably no such explanation can be found.

All in all, Chrisomalis' book is an impressive accomplishment and a valuable contribution to our understanding of the fundamentals of mathematics as a cultural activity.