The Comparative History of Numerical Notation

Stephen Chrisomalis
Department of Anthropology
McGill University

May 2003

A thesis submitted to McGill University
in partial fulfilment of the requirements of the degree of
Doctor of Philosophy

Table of Contents

Abstract

Numerical notation systems are structured, visual, and primarily non-phonetic systems for representing number. This study employs a diachronic and comparative framework to examine over 100 systems used during the past 5000 years. The historical context of each system's origin, transmission, transformation, and decline is traced, linking systems together into phylogenies, but according priority to neither analogical or homological explanations. Structural aspects of numerical notation systems are compared and the limits of variability among them are established. A two-dimensional typology is presented that analyzes the intraexponential and interexponential structuring of each system, in addition to one or more numerical bases. In previous approaches, the only relevant factor considered was the presence or absence of positionality, which led inevitably to unilinear and progressivist conclusions. The analysis of historical relations among numerical notation systems permits a direct approach to questions of how and why they changed. The application of a multilinear cultural evolutionary framework reveals both synchronic and diachronic regularities among numerical notation systems. Where possible, these cross-cultural regularities are related to principles of cognitive psychology. Full explanations of the cultural evolution of numerical notation must also take account of social factors because changes in systems are always the product of decisions made in particular social contexts. Most numerical notation systems are used only for recording and communication, not computation, so it is illegitimate to evaluate their usefulness for functions for which they were never used. A model is presented that relates structural features of numerical notation systems to the contexts of their use and transmission. Because positional systems are most useful for functions related to administrative and scientific institutions that promote cultural hegemony, the observed trend towards positional numerals is a consequence of the dominance of societies that possess such institutions rather than the numerals' inherent superiority.

Résumé

Les systèmes de notation numérique sont des systèmes structurés, visuels et principalement non-phonétiques pour représenter les nombres. Cette étude examine plus de 100 systèmes utilisés au cours des 5000 dernières années en utilisant des techniques diachroniques et comparatives. Le contexte historique de l'origine, de la transmission, de la transformation et du déclin de chaque système est tracé, reliant les systèmes en phylogénies, sans accorder de priorité ni à l'explication analogique ni à l'explication homologique. On compare les aspects structuraux des systèmes de notation numérique tout en établissant les limites de leur variabilité. On présente une typologie bidimensionnelle qui analyse la structure intraexponentielle et interexponentielle de chaque système, en plus d'une ou plusieurs bases numériques. Dans des études précédentes, le seul facteur pris en compte était la présence ou l'absence du principe de position, ce qui a mené inévitablement à des conclusions unilinéaires et progressivistes. L'analyse des rapports historiques entre les systèmes de notation numérique fournit une approche directe afin d'examiner la façon et la raison pour lesquelles ils ont changé. La mise en place d'une théorie multilinéaire de l'évolution culturelle indique des régularités synchroniques et diachroniques parmi les systèmes de notation numérique. Dans la mesure du possible, les régularités découvertes par la comparaison culturelle sont liées aux principes de la psychologie cognitive. Les explications complètes de l'évolution culturelle de la notation numérique doivent également tenir compte des facteurs sociaux parce que les changements des systèmes résultent toujours des décisions prises dans des contextes sociaux spécifiques. La plupart des systèmes de notation numérique ne sont employés que pour l'écriture et la communication des nombres, et non pas pour le calcul, ainsi ce serait illégitime d'évaluer leur utilité pour des fonctions pour lesquelles ils n'ont été jamais employés. On présente un modèle qui relie les caractéristiques structurales des systèmes de notation numérique aux contextes de leur utilisation et transmission. Puisque les systèmes de position sont les plus utiles pour les fonctions liées aux

institutions administratives et scientifiques qui favorisent l'hégémonie culturelle, la tendance qu'on observe vers des systèmes de position résulte de la dominance des sociétés qui privilègient ces institutions plutôt que la supériorité inhérente de ces systèmes.

Acknowledgements

A project of this nature and scope could not have been undertaken without a great deal of assistance and support. First and foremost, I wish to thank my supervisor, Bruce Trigger, for his advice, editorial assistance, financial and moral support, and mentorship during the course of my research, and for promoting innovative methodological and theoretical approaches to anthropological research. I also wish to thank my faculty advisors, Mike Bisson, André Costopoulos, Jim Lambek, and Jérôme Rousseau, for their invaluable advice and assistance and for reading drafts of various parts of this manuscript. Among many others, John Baines, William Clark Bartley, Stephen Houston, Alan Johnston, Antonio Loprieno, Gyula Priskin, and Konrad Tuchscherer helped clarify factual issues in my research and supplied me with additional references and resources. Many thanks to the support staff of McGill University's Department of Anthropology, especially Rose Marie Stano. Many thanks also to the staff of McGill University's interlibrary loans for their tireless assistance in hunting down many obscure references. This dissertation was finished, in part, with funds from the Social Sciences and Humanities Research Council of Canada (Doctoral Fellowship 752-97-1332) and from a McGill University Faculty of Graduate Studies and Research Fellowship. Finally, for her editorial assistance, words of wisdom, and constant support, I dedicate this thesis to my wife Julia, without whom this project could never have been completed.

Chapter 1: Introduction

Our society is one in which number, as expressed through numerical notation, is pervasive. The social and cognitive effects of a concise symbolic means of expressing every number uniquely are inestimable. It is virtually impossible to imagine our industrial civilization functioning without the digits 0 through 9 or a similar system. Nevertheless, we do not often reflect on the circumstances behind the development of such an integral part of our civilization, nor do we realise the extent to which number permeates our lives. We also do not consider the possibility that it may not always have been so. How do other societies, past and present, represent number, if at all? Why does the visual representation of number figure so prominently in complex societies? What cognitive and social functions are served by numerical notation systems in societies that possess them, and why do other societies not possess them?

If you look up from this page and examine your surroundings, I am certain that you will encounter at least one instance of numerical notation in your survey, and probably more. Moreover, unless you have a Roman numeral clock nearby, I am nearly certain that all of the numerals you encounter are those of our own system (the Hindu-Arabic or Western[1]). In our daily lives, numerals serve a wide variety of functions: mnemonic - "Call George, 876-5000", computational - "21.00 x 1.15 = 24.15", valuational – "25 cents", ordinal - "1. Wash dishes, 2. Sweep floor, 3. Finish dissertation", and so on. If we estimate, conservatively, that each of us encounters five instances of numerical notation per minute, sixteen hours per day, we will encounter 1,753,200 instances of

---

[1] The conventional term for our system in popular literature, "Arabic numerals" and the term used in most scholarly literature, "Hindu-Arabic numerals" can lead to considerable confusion given that the scripts used to write the Hindi and Arabic languages use numerical notation systems that, while similar to ours in structure, differ significantly in the shapes of the digits. I use the term "Western" to refer to our system, because the development of our number-forms took place in Western Europe in the late Middle Ages.

numerical notation per year, or, over an average 70-year lifespan, 122,724,000 numerals! Needless to say, most of these will not strike us as being particularly important. Still, it cannot be denied that we encounter far more written numbers in our lifetime than we do sunsets, songs, or smiles. This fact is even more notable because of the likelihood that, until the past few centuries, the opposite was certainly the case for most individuals.

These digits are so prevalent that many people today equate our numeral-signs with the set of abstract numbers. For them, 62 does not merely *signify* the abstract concept "sixty-two" - it is in fact the purest form of the number itself, the stuff of pure mathematics (or perhaps pure numerology). The fact that these signs are encountered and used within the context of formal mathematics doubtless contributes to the prevalence of such attitudes. While many people in the West are familiar with Roman numerals, most people regard them only as an archaic curiosity, a hopelessly inefficient vestige of a long-dead civilization. According to this view, our numeral-signs constitute abstract number, and other systems (when recognized as such) are simply deviations from the Platonic entity comprised by these signs.

This view is entirely erroneous, and rests on the confusion of a mental concept (signified) with its symbolic representation (signifier). Our numerical notation system has an extensive history, as do the more than one hundred systems that have existed over the past five thousand years. Still, the worldwide prevalence of Western numerical notation is undeniable, and this has led many scholars to assert its supremacy solely on the evidence of its near-universality (Guitel 1975; Ifrah 1985, 1998; Dehaene 1997; Zhang and Norman 1995). Despite the efforts of the so-called "ethnomathematics" movement to recognize the mathematical achievements of non-Western civilizations (Joseph 1991; Ascher 1991), the study of numerical notation remains mired in a theoretical framework that has much more in common with late-19th century unilinear evolutionism in anthropology than it does with early-21st century critiques of unfettered scientific and moral progress.

Various false beliefs about the perpetuity and superiority of our own numerical notation system are intertwined with assumptions regarding the history of numerical notation. To be sure, the numerals 0 through 9 are used extraordinarily widely, so much so that virtually all literate individuals worldwide, as well as a very sizable number of illiterates, understand them. Nor is there any competing system with any reasonable chance of supplanting our system in the near future. Nevertheless, this fact does not imply that our system will dominate the whole world forever. Yet many who would never think of accepting Fukuyama's (1992) assertions regarding the end of economic history have no qualms about regarding the eventual universality of Western numerical notation as inevitable and natural. Such unilinear and progressivist assertions are anathema to the anthropologist and the historian of science. If in fact the Western system is superior (which raises the question of what we mean by superiority), then we ought certainly be able to find evidence of its superiority, rather than simply making unfounded assertions.

Numerical notation as a topic of academic study is, in fact, a relatively common pursuit, with linguists, epigraphers, archaeologists, anthropologists, historians, psychologists, and mathematicians all making significant contributions to the literature. However, these studies are for the most part restricted to the analysis of one or a few numerical notation systems. A small number of synthetic and comparative works dealing with numerical notation have been published (Cajori 1928-9; Menninger 1969; Guitel 1975; Ifrah 1998). However, there is an almost tragic failure on the part of such scholars to deal with more obscure numerical notation systems, such as those of sub-Saharan Africa, North America, and Central Asia. While this is partly due to recently published discoveries of new systems, not all these omissions can be explained in such a manner. Even in the synthetic studies that have been undertaken, little effort has been made to fit the descriptive findings relating to each numerical notation system into any theoretical framework, be it anthropological, historical, or psychological. It is this failing, rather than the lack of factual evidence, that has resulted in the current theoretical stagnation of studies of numerical notation.

The present study is a comparative analysis of all numerical notation systems known to have ever existed throughout human history. I will discuss approximately one hundred distinct cases, grouped into eight distinct cultural phylogenies, while treating another fifty or so as minor variants without the need for extensive discussion. By presenting a universal study of such systems and paying attention to the historical connections and contexts in which they are encountered, I hope to present a satisfactory framework that both accounts for cultural universals and evolutionary regularities and remains cognisant of idiosyncratic features. I will distinguish several important types of numerical notation, evaluate them in terms of their efficiency for performing specific functions, and, where possible, relate their features to panhuman cognitive abilities and tendencies.

Definitions

A **numerical notation system** is a visual but primarily non-phonetic structured system for representing numbers. Signs such as 9 and 68, IX and LXVIII, are part of numerical notation systems, but numeral words such as *nine* and *achtundsechzig* are not. Though there are ties between numeral words and numerical notation, a **lexical numeral system,** or the sequence of numeral words in a language (whether written or spoken), has a language-specific phonetic component and often has irregularities that are absent from numerical notation systems. Every language has a lexical numeral system of some sort, whereas numerical notation is an invented technology that may or may not be present in any given culture.[2] Some numerical notation systems contain a small phonetic

[2] However, see Hurford 1987: 68-78 for arguments contrary to the assumption of the universality of lexical numeral systems.

component, as in *acrophonic* systems where number-signs are derived from the first letter of the appropriate number-words in a given language. However, since such systems are still understandable without having to understand a specific language, they are numerical notation systems.

Numerical notation systems must be *structured.* Simple and relatively unstructured techniques, such as marking lines on a jailhouse cell to count one's days or piling pebbles in a basket, are largely or entirely unstructured, excluding them from the sort of analysis that can be undertaken of more complex systems. These techniques have in common a reliance on **one-to-one correspondence**, in which things are counted by associating them with an equal number of marks or other identical objects. On the other hand, a numerical notation system, is a set of different **numeral-signs**: single elementary symbols used in combination to represent any given number (e.g. X, ∩, [illegible], [illegible], 6).[3] A **numeral-phrase** is a group of one or more numeral-signs used to express a given number (e.g. MMDXXV); the exact number of signs in a numeral-phrase depends on the number being expressed and the structure of the numerical notation system. All numerical notation systems (and most numeral systems) are structured by means of **exponents** of one or more **bases.** The term **exponent** refers to the number of times that a number X is multiplied by itself; $10^1=10$, $10^2=100$, $10^3=1000$, etc. Any number raised to the exponent 0 equals 1. A **base** is a natural number B in which exponents of B are specially designated. While mathematicians normally require that a base be extendable to an infinite number of exponents of B (e.g., 10, 100, 1000, 10,000, ... *ad infinitum*), most numerical notation systems are not infinitely extendable. For my purposes, it is sufficient that *some* exponents of B are specially designated within a numerical notation system. Our own and many other numerical notation systems use a base of 10, but this is by no means

---

[3] A few numeral-signs are more complex in that they combine two or more signs into one in order to represent multiplication, e.g. Sumerian [illegible] (=3600x10 = 36,000) and Greek acrophonic 𐅇 (= 5x1000 = 5000), but they are treated as elementary numeral-signs since their use is identical to that of all other simple signs in the systems in question.

universal. Nor is the base of a numerical notation system necessarily (or, indeed, frequently) identical to that of the numeral system used in cultures where it is found. In addition to its base, a numerical notation system may have one or more **sub-bases** that structure it. The Roman numeral system has a primary base of 10 with a sub-base of 5. Unlike bases, the exponents of sub-bases are not specially designated; there are no special Roman numerals for 25 or 125. It is, rather, the *products* of a sub-base and the exponents of the primary base that are specially designated - for the Roman numerals, 50 (5x10) and 500 (5x100).

Two topics that I will study only peripherally, though they are often confused with the study of numerical notation, are **number** and **mathematics**. Number is an abstract concept used to designate quantity. The ontological status of numbers is much in debate among mathematicians and philosophers, but for the purposes of my study, a simple (if philosophically naïve) definition is probably best. Questions such as whether numbers are "real" or Platonic entities, or whether one plus one is really equal to two, are beyond the scope of this study. Similarly, in defining mathematics as the science that deals with the logic of quantity, shape, and arrangement, I am consciously employing a simple definition for a highly contested term. In order to understand numerical notation, one need not have any mathematical ability save knowledge of basic arithmetic. While some parts of mathematics make frequent use of numbers (number theory being the most obvious example), large parts of the discipline have only infrequent or peripheral encounters with numerical notation. Mathematics is not specific to any one numerical notation system, although some systems are more efficient than others for mathematical tasks. Because it is an error so frequently propagated by historians of mathematics, it is important to note that numerical notation systems are not necessarily designed with mathematical purposes in mind. Even in the modern West, where mathematical ability is more extensive than in any other society at any point in history, the function of numerical notation is primarily non-mathematical. This is a point that warrants frequent repetition, insofar as it heavily influences how we interpret the history of numerical notation.

Methodology

Throughout the twentieth century, one of the central dichotomies in anthropological theory has been that of universal cross-cultural explanation versus particularism and relativism. Universal explanation, as typified by the work of E.B. Tylor and later by Leslie White and Julian Steward, has been characterized as the search for "cultural laws" or, at the very least, cross-cultural regularities (Tylor 1958 [1871]; White 1949, 1959; Steward 1955). It was believed that when these laws were discovered, they would provide insights into human nature comparable to those provided by, for instance, psychology, or even, in ideal circumstances, offer analytical tools as powerful as laws in the physical sciences. This perspective has been caricatured as a hopelessly positivistic and simplistic way of analysing anthropological data, and its findings rejected as gross oversimplifications of complex cultural contexts. On the other hand, we are faced with particularism, which maintains that all anthropological data should be interpreted within a framework that emphasizes their historical and cultural context, and regards the search for cross-cultural regularities as misleading. In American anthropology, this tradition extends back to Franz Boas and Robert Lowie, and is represented in modern works by Clifford Geertz and Marshall Sahlins (Lowie 1920; Boas 1940; Sahlins 1976; Geertz 1984). At minimum, this school of thought requires that we carefully consider the social and historical context of anthropological facts in order to take account of the full complexity of human systems – an assertion with which no social scientist would seriously disagree. At its most extreme, it is mocked by its opponents as an exercise in collecting data without any attempt to produce mid- to high-level theoretical statements, or in fact as denying that statements comparing and generalizing data can be made. Obviously, these two extremes are exaggerations. Most comparativist anthropologists and archaeologists undertake detailed field studies of a single group or site and recognize the value of particularizing research, while relativists, in the very act of discussing theories of symbolism and power, are making statements of general applicability to human societies and thus engage in *de facto* cultural comparison.

In the present study, I am attempting to reconcile these two positions, at least in part, through the methodology of *universal diachronic comparison.* By undertaking detailed histories of specific numerical notation systems and then organizing them into cultural phylogenies, I hope to show the complex historical situations under which numerical notation systems develop, persist, change, diffuse, and decline. At the same time, I intend to study any regular patterns of change, or sociocultural evolution, that may emerge from such situations. The problem is partly one of scale; what may seem hopelessly complex at the level of historical detail may be abstracted to a set of more general principles at a larger scale. While generalization necessarily implies some degree of simplification, it is necessary in order to gain a holistic understanding of a phenomenon.

Universal Comparison

This study is a universal one, in that I have not excluded any numerical notation system intentionally for any reason save where data are not plentiful enough to undertake a reasonable analysis. Despite dozens of works on the history of mathematics detailing important numerical notation systems, those who study these systems remain largely unaware of the utilitarian constraints governing their development and use. In part, this is because the total observable variability among numerical notation systems (or any phenomenon) cannot be understood by studying only a fraction of that phenomenon. To paraphrase the old fable, if we study only the elephant's trunk or tail, we ignore most of the animal. Obviously, cases have been omitted where data were insufficient or, in the most extreme instance, where there is no extant evidence concerning a numerical notation system. While this study can be subsumed under the rubric of comparative research, it should be noted that its unit of analysis is not the culture but rather the numerical notation system. Identifying the 'skin of culture' is an interesting if ultimately somewhat misleading ontological problem, but it has no place in this study (Kroeber 1953; Naroll 1964). A single system can transcend the 'boundaries' of culture, language, and script, while still being considered a single entity.

While comparative research aims to produce universal generalizations about human behaviour, using the universe of cases rather than a sample is neither possible nor desirable in most cross-cultural studies. In many instances, it would be extremely difficult to examine the entire universe of relevant cases. Even if this were possible, in order to use analytical statistics on quantified cross-cultural data, each case must be independent from the others. In cross-cultural research, this requires that each feature in a sample may not be historically derived or diffused from any other feature. This issue, known as *Galton's problem*, is the thorniest methodological issue in statistical cross-cultural research. Efforts to negate its effect in cross-cultural research have occupied statistically minded anthropologists for decades. Yet doing so effectively excludes the investigation of two very important issues - change in a system over time and the development of new systems based on existing ones - both of which involve the dependence of one case or another.

While in a statistical study, devoted to establishing correlations between two phenomena (for instance, population size and social inequality), sampling is absolutely necessary in order to ensure that homologies are excluded, the present study presupposes that homologies should not be excluded. Rather, historical connections between two systems are important both as historical data in their own right and in order to examine the differences between analogies and homologies. As one of my goals is to establish cultural phylogenies of numerical notation systems, all systems, regardless of their historical relatedness to others, must be included. Indeed, there are many situations where homologies may be of greater importance than analogies for studying evolutionary regularities or universals. There may in fact be enormous variety among numerical notation systems that are historically connected to one another; to omit all but one is extraordinarily problematic. I am not arguing that statistical methods have no place in cross-cultural research. The establishment of correlations between two traits on a worldwide or regional basis is of enormous utility for researchers. However, we must recognize the limitations of this procedure for answering a wide variety of questions - not only those of particularistic interest (diffusion and historical connections), but also ones of interest to the student of cross-cultural regularities and universals, such as the total worldwide variability of a given phenomenon.

The methodological insistence on the independence of cases becomes more troubling when it leads to the assumption that independent invention of cultural traits is the result of functional considerations, while diffusion is "merely" particularistic and historical. In this assumption, we see the re-emergence of the unproductive dichotomy between universalism and particularism. My examination of numerical notation demonstrates that diffusion is, as Julian Steward (1955) noted, in many ways similar to independent invention, and that the circumstances under which a system diffuses often involve significant considerations of function and utility. It is frequently forgotten that the transmission of a phenomenon from one society to another is a complex process of communication, resistance, trial, and adoption (or rejection), and that such processes are inventive, not merely receptive. It is entirely incorrect to imagine that simply because an innovation had an antecedent, it ought not to be accorded the same theoretical significance as a truly "independent" invention - presuming, of course, that a truly antecedentless invention is not merely a theoretical fantasy developed to satisfy artificial standards of methodological rigour. It may be that historically particular cultural features (such as a language's lexical numerals) are important in determining the form and structure of independently invented numerical notation systems.

Of course, in failing to exclude numerical notation systems derived in part from other systems, I am unable to undertake any statistical analysis, save that of basic descriptive statistics. However, as there are only seven probable instances of the independent invention of numerical notation systems, such an analysis would not be possible in any case due to the limited universe that could be examined. As all of these independent cases are included herein, interested readers may conduct such an analysis themselves, should they judge my comparison of historically related cases unacceptable.

Culture History and Diachrony

One of my major goals in this study, perhaps the most important one, is the reconstruction of the cultural history of numerical notation. The prestige accorded to culture history by anthropologists and archaeologists has declined greatly over the past half-century. The term itself carries archaic connotations reminiscent of cataloguing pictorial symbols or arrowheads without any theoretical content. As Trigger (1989: 122-124) notes, the culture-historical school in early and mid-20th century anthropology and archaeology was more interested in discerning the synchronic spatial patterns of decontextualised traits than in studying diachronic or chronological sequences of cultural phenomena. I am pursuing this latter goal when I refer to culture history: the compilation of data into historical sequences using documentary and archaeological evidence (cf. Mace and Pagel 1994). I intend to restore a truly diachronic perspective to its proper place in the study of cultural history by reconstructing the histories of past and present numerical notation systems.

While cultural history is interesting in its own right, if it were the case that the history of numerical notation was well understood, then perhaps the exercise would be pointless. However, this is not the case. A large number of the cases I have studied have not been mentioned in any previous synthetic study of numerical notation - for example, the Pahawh Hmong, Berber, Bambara, Kitan, Cherokee, and a host of others. Moreover, these systems are not simply variant forms of other more commonly known systems; each possesses unique structural features and their omission has seriously weakened the universal validity of others' work. It is noteworthy that these and other understudied systems hail from peripheral societies whose creativity has often been denigrated at the expense of cultural cores. Even well known systems are in some cases quite poorly understood. For instance, no publication apart from specialist works on classical antiquity has ever discussed the date and circumstances of the initial use of Roman numerals and no general textbook on the history of mathematics or popular work on the history of numerals tackles the issue. To disregard culture history as mere grubbing for facts is to disregard an important part of the task of historical analysis.

Culture history, as I will be using the term, encompasses both the study of each numerical notation system as a historical entity in its own right and the study of the historical correlations among different systems. Recognizing that the definition of a numerical notation system as a unit in time and space is a tricky one, one must endeavour to distinguish among individual systems in the least arbitrary way possible. For the purpose of this study, any change in the structure of a system is a necessary and sufficient condition to treat the antecedent and successor forms as separate cases. For instance, because the classical Roman numerals have a somewhat different structure from the medieval Roman numerals, they are treated separately. However, for the most part, simple changes in the shapes of numeral-signs without a concomitant structural change are insufficient to warrant separate treatment of different cases. Thus, I treat many of the South Asian systems together since they are structurally identical to their predecessor, the Indian positional system. In a few cases, the amount of data available on a given system or an enormous geographical and cultural separation of antecedent and successor systems necessitates treating them separately even though they are not structurally different. This need is particularly acute where the systems in question are historically very important, geographically widespread, and/or have a large number of successor forms. I treat the structurally identical Indian, Arabic, and Western positional systems, which are directly related to one another, as separate cases because each is used over a long period of history in many different societies. After identifying and studying each distinct case, I have placed systems in historical relation to one another, constructing cultural phylogenies of numerical notation systems.

Diachronic research is not simply a matter of determining *when* a system existed, nor is diachrony synonymous with history. Rather, it is an active means of establishing changes in systems and of correlating those changes with changes occurring through time in other societal sub-systems. Diachronic research with respect to numerical notation comprises at least five different aspects. These are: a) the circumstances under which a system came originally to be developed, whether through independent invention or through transmission with variation; b) when, how, and for how long a system came to be adopted by a society or some portion thereof; c) whether any changes in the system occurred throughout its use, and if not, the conditions under which it persisted through time; d) whether, one or more times during its existence, the system spread from its initial locus of use to other societies, with or without some modification occurring upon transmission; and e) if the system is no longer in use, the time and circumstances under which it ceased to be used.

Diachronic research satisfies two primary research goals. First, it enables the construction of cultural phylogenies of numerical notation systems. While this task is more descriptive than analytical, it is a necessary step in distinguishing homologies from analogies, as well as being of historical interest. Second, it facilitates the examination of patterns of change and cultural processes, not only within a given society but also among a set of related societies. The purpose is not merely to place phenomena in historical perspective (though this is certainly an important goal). The study of sociocultural evolution must be diachronic if it is to depict worldwide regularities in patterns of change. It also requires, as much as possible, that the same questions be answerable for each system under consideration, so that comparisons will be rigorous.

Most comparative research is synchronic, examining two or more traits within a number of societies at a given time. Many cross-cultural studies use historical delineation, or the indication of when in time a given phenomenon existed in a particular society, in order to permit a trait or phenomenon to be historically contextualized. While this is useful, pinpointing the historical context of a given society in cross-cultural research is one-dimensional with respect to time - a point, rather than a line. In order for a study to be diachronic, it must examine the same society, trait, or institution using at least two chronological points, and, ideally, along the entire timeline between these two points. Because most cross-cultural studies ask functional questions, which seek to establish how one trait is related to another within a social system at a given time, rather than historical ones, which ask how traits and systems change through time, it is perfectly understandable that diachronic cross-cultural research would not be an appealing methodology.

However, researchers often use synchronic data alone to draw diachronic conclusions using theories that predict the patterns of change one might expect in the cases under study. Inferential reconstruction, in one sense, is pervasive in historical disciplines; history and archaeology can only reconstruct the past inferentially. However, what I mean by inferential techniques is not simply that the past is reconstructed from existing evidence, but rather that general theoretical principles such as sociocultural evolution, distributional patterns or diffusion are invoked to infer history from existing synchronic patterns. This research strategy has proven effective in the reconstruction of regional culture histories using existing distribution patterns of linguistic and anthropological features (Jorgensen 1974; Driver 1966). If the assumptions made in an inferential study are valid and if there is some way to test the reconstructions after their completion, diachronic inferences are perfectly legitimate. For instance, Aberle (1974) argues that it may be possible for ethnologists to reconstruct cultural histories by employing principles derived from historical linguistics, a discipline with a long tradition of reconstructing diachronic patterns of change. Other inferential techniques, such as the "age-area" principle, that widely distributed traits are likely to be older than associated traits of less widespread distribution, are now quite rightly discounted by anthropologists (Wallis 1925). The use of cladistics to reconstruct the histories of genetic populations or language families based on clusters of shared traits is an important inferential tool for hypothesis formulation regarding past events. However, cladistic analysis of cultural phenomena (including language) is problematic due to the frequency of inter-societal borrowings and the required assumption of the monogenesis of the phenomenon. In any case, such techniques are not easily transferred to numerical notation systems, since they lack a quantifiable unit of comparison (such as the gene or the cognate) that can be analyzed statistically. Because sufficient data exist for the direct reconstruction of phylogenies of numerical notation systems, cladistics is a superfluous and less reliable technique for analysing numerical notation.

Wherever possible, I (and most historically-minded anthropologists) prefer to use *processual*[4] data in the reconstruction of diachronic patterns. Processual analyses examine data from many different periods and seek to account for the resulting sequences in terms of some mechanism by which culture change took place. Of course, the extent to which a particular analysis uses inferential and processual data most often depends on the sort of data available, with inferential data used to fill in the gaps left in the historical record. While inferential reconstructions are primarily deductive, processual analysis relies much more heavily on induction, though of course all explanations of the

---

[4] My use of the term 'processual' should not be confused with processual archaeology, which rarely analyses cultural change in the manner I describe.

mechanism cultural process will be theory-laden - for instance, whether human agency, mental structures, sociopolitical conditions, or environmental factors played the role of 'prime mover' in determining patterns of change.

An important consequence of my use of diachronic methodology is that, although this study is motivated by questions of anthropological theory, the main sources of data that I must rely on are historical rather than anthropological. Archaeological data will be of some use (particularly for the New World), but even though archaeology is a discipline whose emphasis is clearly diachronic, its value is limited by the fact that numerical notation is primarily known through written documents that are only infrequently considered by prehistoric archaeologists. History and its auxiliary sub-disciplines, especially epigraphy and palaeography, are indispensable tools for the study of numerical notation. The work of linguists interested in writing systems is also an abundant source of data. In many cases, data that are apparently only of historical interest within a small field may come to be of enormous theoretical significance when placed in the context of cross-cultural anthropology.

Despite my reliance on historical data, the history of numerical notation is not typical history by any means, not even in relation to the history of science and technology. Our lack of evidence concerning those who invented and adopted particular numerical notation systems is almost total; only in a few instances do we have specific knowledge of individual innovators and adopters of systems. In some cases, all that remains of a system are a few hastily inscribed graffiti, or a small number of untranslated documents. As a result, although numerical notation is obviously created and used by human beings, the data necessarily obscure the human aspect of these systems. It is my goal to reconstruct the processes of creation, transmission, and use of such systems despite the absence of such evidence.

Because this study covers over five thousand years of history and spans several continents and at least half a dozen academic disciplines, the data used have not been collected by me first-hand but have been compiled from existing published records. I have been unable to use the vast scholarly material written in Arabic, Chinese, and Japanese because of my lack of proficiency in those languages. In any study of this scope, the researcher is bound by the topics that others have chosen to study. As a result, valuable information may remain unpublished at present. I hope that I have done justice to the work of others in bringing their data into a comparative anthropological framework.

Structural Typology of Numerical Notation

The systematic classification of numerical notation systems is essential for determining the relevant features of systems, for distinguishing independent inventions of those features, and for determining their utility for various functions. I will demonstrate that past attempts have failed to classify existing systems adequately and in a way that is analytically useful. The goal of typology is not simply to develop a scheme into which every case will fit, but to do so in a way that allows an analyst to ask and answer questions of the data that could not otherwise be considered. When poorly done, typology is descriptive but non-analytical, and thus essentially useless; when well done, it organizes knowledge in a way that answers enquiries.

Any classificatory scheme is inherently theory-laden, and serves only to answer a limited variety of questions that might be asked of a set of data. Yet I believe that the typology I present here represents all the principles by which numbers are represented and emphasizes the features of numerical notation that are cognitively most important. The questions it helps to answer are not those of culture history or chronology; these issues are best resolved through diachronic comparison and the construction of cultural phylogenies. Rather, my structural typology removes each system from its temporal, geographic, and spatial contexts and considers only the way in which numeral-signs are combined to represent any given number. In so doing, I am able to show that there are only five basic types of numerical notation, and furthermore that these are formed using only two general principles.

Geneviève Guitel's *Histoire comparée des numérations écrites* (1975) remains the most comprehensive scholarly work devoted to numerical notation. As Guitel was a historian of mathematics, over half of her text is devoted to mathematical, calendric, and other phenomena that do not deal directly with numerical notation, and those parts that do discuss numerical notation contain virtually no historical data or social context. Furthermore, her failure to analyse systems she considers 'primitive', such as the Phoenician, Minoan-Mycenean, and Inka systems, seriously limits the significance of her analysis. The one lasting contribution of her work is its attempt to classify numerical notation systems according to their structure.[5] Guitel's effort is the only systematic attempt to catalogue and classify numerical notation systems.[6]

For each system under consideration, she codes units or primitives such as 4 "*s*" (= *symbolisé*); numbers represented by juxtaposition through addition (e.g. III) "*j*" (= *juxtaposé*); those where multiplication is understood as in number words such as *fifty* "*l*" (= *ligaturé*); and those formed with the use of the zero (400) "Σ". It is thus possible, she claims, to describe in full any number in any written numeration system (a claim rendered true only by ignoring or using special symbols for systems that do not follow simple rules, as the Hindu-Arabic system does). She then classifies approximately twenty-five systems (drawn from about a dozen societies) according to which of her operations they use. She classifies systems into Type I (use of additive symbols alone); Type II (hybrid systems using both addition and multiplication in various combinations),

---

[5] The works of Georges Ifrah (1985, 1998), which are popular in style and much more widely read (and translated) than Guitel's, implicitly follow her typology throughout. This has led to its adoption in many recent histories of mathematics.

[6] Zhang and Norman's (1995) cognitive taxonomy, discussed below, is oriented towards description for the purposes of evaluating efficiency and thus has limited use as a general classificatory scheme.

and Type III (positional notation systems such as our own). Each of these is subdivided according to the specific way in which her operations are used to form numeral-phrases and according to the numerical base(s) of the system. Some examples of Guitel's classifications are shown in Table 1.1.

Table 1.1: Guitel's typology of numerical notation systems

System	Classification	Sample
Roman	$I_{A''}$	678 = DCLXXViii
Greek alphabetic	$I_{C''}$	678 = $\overline{\chi o \eta}$
Akkadian	II_A	678 = 
"Greece III" alphabetic	II_B	678 = $\overline{\chi o \eta}$ (40,678 = $M\bar{\delta}.\overline{\chi o \eta}$)
Babylonian	$III_{A'}$	678 = 
Western	III_B	678

Despite an admirable attempt, Guitel's analysis fails the most basic test of classification, which is that it must classify similar systems together and separate dissimilar ones. While the Greek alphabetic system is identical to the "Greece III" system for all numbers under 10,000 and differs only in the addition of a special multiplier-sign for 10,000, merely on account of this fact the former is placed in Group I and the latter in Group II. Similarly, the Akkadian and Babylonian systems are identical for all numbers below 100, and are based on the principle that a repeated numeral-sign within an exponent implies that the signs should be added, but Akkadian ends up in Group II and Babylonian in Group III. Moreover, comparing the two systems in each of Groups I, II, and III listed above, it is evident that, while they share some features in common, they are by no means similar, certainly not *more* similar to one another than they are to other systems. The Roman system uses nine numeral-signs drawn from a total of seven different sign types to express what in the Greek alphabetic system takes only three signs drawn from twenty-seven different types, yet both are Group I systems.

The essential problem with Guitel's scheme is that its primary division is made on the basis of only one factor: the degree to which multiplication is used in forming a given numeral-phrase. Her system is one that assumes that, since the invention of positional notation was very important in the development of our numerical notation system, positionality is therefore the primary criterion by which all numerical notation systems should be judged. Because it only considers to what extent multiplication helps determine the value of a numeral-phrase, it ignores several major features of numerical notation systems (perhaps more important than position, as I shall argue below). Furthermore, it is inherently tied to the questionable claim that the primary function of numerical notation is mathematical; since positional notation is used in modern mathematics and arithmetic, it is presumed to be the benchmark against which all other systems ought to be evaluated. Guitel's system is not factually incorrect in any significant way; systems that are identically structured do end up in the same category. Yet because it is tied to the misleading question "To what extent does system X use the principles of multiplication and positionality?", it fails to represent fully the similarities and differences among numerical notation systems.

In contrast to Guitel's analysis, in which primary distinctions are made only on the basis of positionality, I believe we must consider two separate dimensions of numerical notation systems in order to classify and analyse them adequately. I call these dimensions **intraexponential** and **interexponential** organization. Each of these dimensions is further subdivided as described below. These rather complex descriptions are followed by a number of examples of the various types of system, which I hope will clarify matters.

Intraexponential organization determines how numeral-signs are constituted and combined within each exponent of the base. The major types of intraexponential organization are **cumulative, ciphered,** and **multiplicative**. Cumulative systems are those in which the value of any exponent of the base is represented through the repetition of numeral-signs, each of which represents one times the exponent value of the sign, and which are then added; for instance, XXX = 30 in the Roman system because the sign for 10 (X) is repeated three times. Ciphered systems, on the other hand, use at most a single numeral-sign for each exponent represented, with different signs being used to represent different multiples of the exponent. Our own (Western) system is a ciphered one: a number that has as 10^3 (thousands) as its highest exponent (e.g. 1984) will require at most four symbols, one for each exponent of the base. Multiplicative systems are those in which there are two components for each exponent represented: a unit-sign (or sometimes multiple signs), which represents the quantity of that exponent needed to represent the number, and an exponent-sign, which represents an exponent of the base of the system. The product of the two signs is then taken to determine the value of that exponent. The multiplicative principle is rarely used throughout an entire numerical notation system, but is often only used to structure higher exponents of the base.

Interexponential structuring determines the way in which the values of the signs for each exponent of the base are combined to symbolize the value of each entire numeral-phrase. Interexponential organization is sub-divided into **additive** and **positional** sub-types. Additive systems are those in which the values of each exponent represented in a numeral-phrase can simply be added up to produce its total value. For instance, the Roman numeral CCLXXVIII consists of two 100s (10^2), one 50 (5×10^1), two 10s (10^1), one 5 (5×10^0), and three 1s (10^0), for a total of 278. Positional systems, of which the Western system is the best known, are those in which the value represented within a given exponent is determined not only by its constituent numeral-signs but also by its position within the numeral-phrase. The intraexponential values within a numeral-phrase must all be multiplied by the appropriate exponent-values before the sum of the phrase can be taken. Within the Western numeral 90642, 9 represents (9×10^4), followed by (0×10^3), (6×10^2), (4×10^1), and (2×10^0). In a positional system, it is necessary that exponents be listed in order, because the positional ordering of signs determines their value. Interestingly, although there is no logical requirement that additive systems list their exponents in order, they almost universally do so. The Roman numeral CCLXXVIII could be unambiguously read even if it were written as VIIICCXXL, or even XLIVIXCIC, if we omit the slight complexity brought about by the occasional use of subtraction; yet both additive and positional systems are strictly ordered interexponentially.

All numerical notation systems are structured both intra- and interexponentially, creating a total of six theoretically possible pairings of principles. However, it is logically impossible for a multiplicative-positional system to exist because multiplicative systems represent the required positional value (10, 100, 1000, etc.) within each exponent, leaving only five possibilities. The five basic possibilities are detailed in Table 1.2.

Table 1.2: Typology of numerical notation systems

	Additive The sum of the values of each exponent is taken to obtain the total value of the numeral-phrase.	**Positional** The value of each exponent must be multiplied by a value dependent on its position before taking the sum of the numeral-phrase.
Cumulative Many signs per exponent of the base, which are added to obtain the total value of that exponent.	Classical Roman 1434= MCCCCXXXIIII (1000+4x100+3x10+4x1)	Astronomical cuneiform 1434 = 𒎙𒐈 𒐐𒐘 (2x10+3)x60 +(5x10+4)
Ciphered Only one sign per exponent of the base, which alone represents the total value of that exponent.	Greek alphabetic 1434= ͵α$\bar{\upsilon}\bar{\lambda}\bar{\delta}$ (1000+400+30+4)	Bengali 1434 = ১৪৩৪ (1x1000+4x100+3x10+4x1)
Multiplicative Two components per exponent, unit-sign(s) and an exponent-sign, multiplied together, give that exponent's total value.	Chinese (traditional) 1434=一千四百三十四 (1x1000+4x100+3x10+4)	LOGICALLY EXCLUDED

Cumulative-additive systems, such as Roman numerals, have one sign for each exponent of the base; the signs within each exponent are repeated and their values added, and then the total value of all of the signs can be taken by summing all the numeral-signs in the phrase. **Cumulative-positional** systems likewise use repeated signs to indicate the value of each exponent, but this value is then multiplied by the appropriate exponent values (in the Babylonian example above, 60 and 1) before summing the phrase. In order to be entirely unambiguous, some sort of placeholder or zero sign is required. **Ciphered-additive** systems have a unique sign for each multiple of each exponent of the base (1-9, 10-90, 100-900, etc. in a base-10 system like the Greek alphabetic); the values of these signs are added to obtain the result. **Ciphered-positional** systems such as our own have unique unit signs from one up to but not including the base (e.g. 1,2,3...9) and a zero sign; the unit-value is multiplied by the exponent-value indicated by its position, and then the sum of all the values gives the total value of the numeral-phrase. Finally, **multiplicative-additive** systems (like the traditional Chinese system shown above, but also for that matter spoken English numeral words; e.g. *three thousand six hundred and twenty four*) juxtapose a unit-sign (or signs) and an exponent-sign, which are multiplied together, and then the sum of those products gives the total value of the phrase.

Most numerical notation systems use only one of these five organizational combinations throughout the entire system. However, a number of additive systems use one intraexponential principle (either cumulative or ciphered) for lower exponents of the base, and then use the multiplicative principle thereafter. These systems comprise about 30% of those I examine in this study. Systems that use two principles are not exceptions to my typology. They simply need to be analyzed in two parts, with each part being assigned the appropriate principle. For instance, the version of the Greek alphabetic system shown in Table 1.2 is ciphered-additive, but for the thousands exponent and higher, it is multiplicative-additive. Thus, it is ciphered-additive for exponents below

1000 and multiplicative-additive for those above 1000. This feature is the multiplicative component that leads Guitel to place this version of the Greek alphabetic numerals in her Type II as opposed to Type I. Contrarily, I classify the system that lacks multiplication as ciphered-additive and its partly multiplicative counterpart as ciphered-additive with multiplicative-additive structuring for exponents above 1000, thus delineating the similarities and differences between these variants. No numerical notation systems employ more than two of the five basic types, and no positional system uses more than one type.

Systems that have a sub-base as well as a base require further typological clarification because they can use two intraexponential principles: one for units up to the sub-base, and others for multiples of the sub-base up to the base. For instance, the cumulative-positional Babylonian system shown in Table 1.2 has a base of 60 and a sub-base of 10. In this case, we must know both how units from 1 to 9 are expressed and how tens from 10 to 50 are expressed in order to fully describe its intraexponential structure. In this case, both the sub-base and the base use the cumulative principle, so we might more properly describe this system as a (cumulative-cumulative)-positional system. However, very few systems use a different intraexponential principle for their sub-base than for their base, so this elaboration is mostly unnecessary. I will only use this more complex terminology when a system uses two different intraexponential principles. Again, none of this affects the *inter*exponential structure of these systems.

It is my belief that this typology better reflects the different features of numerical notation systems than does Guitel's, in particular because it reflects both intra- and interexponential principles. It shifts the focus of analysis from *systems* to the structural *principles* that are used to build systems, and thus allows a more nuanced examination of systems' structure. Furthermore, it allows us to ask fruitful questions regarding the evolution and historical connections of different systems. Even if it were the case that the primary vector in the evolution of numerical notation is a unilinear progression from the use of addition to the positional use of multiplication, we lose a great deal of information about these systems by ignoring intraexponential considerations.

The Ontology of Number and the Comparative Method

Throughout this study, I treat numerical notation as a symbolic system that is almost completely translatable cross-culturally without any loss of information or change of meaning. I regard the number 1138 as essentially identical in meaning to MCXXXVIII or 𓆼𓍢𓎆𓎆𓎆𓏺𓏺𓏺𓏺𓏺𓏺𓏺𓏺 or any other representation. Of course, the ways in which these systems are structured differ greatly. Moreover, there may be a whole host of social and symbolic attachments to any system. Nevertheless, the core of the system is the representation of number, and number, as a logical concept, is one of very few unassailable universals in the sphere of human knowledge. In semiotic terms, although the linguistic and symbolic signifiers for numbers may differ greatly (23, *dreiundzwanzig*, XXIII, *viginti tres*, etc.), the signified is in each case identical. There can be no "errors in translation" in the representation of numbers, at least in their basic numerical meaning. Of course, the symbolism attached to numbers (such as in numerology) varies quite widely in different societies, and new numbers outside the set of natural numbers can be developed as necessary, particularly within the context of formal mathematics. Nevertheless, the correlation of both numeral-signs and lexical numerals with the set of natural numbers is not culturally relative.

This does not imply that numerical notation is inherently iconic, which means that the signifier is actually a visual representation of the signified (as in pictorial signs for gendered washrooms, for instance). Instead, as with traffic lights, the relationship between signifier and signified is arbitrary. The association of single lines or dots with the units (e.g. the Roman numeral III) and the resulting cumulative intraexponential structure of many numerical notation systems are somewhat iconic. It seems likely that the cognitive principle of one-to-one correspondence between units and single strokes or dots reflects the phenomenological perception of quantity. However, other theories of iconic representation in numeral-signs, such as that which holds that the Roman numeral V represents a spread hand of five fingers and that X represents two hands placed together (10 total fingers), are today considered highly dubious (Keyser 1988). There is nothing universal about the signs 0 through 9 that explains their worldwide universality, or that suggests that the meanings of any two of these symbols could not be exchanged.

Most mathematicians treat abstract numbers as if they were Platonic entities, having a reality external to human beings despite not having a physical reality. We need not go so far in order to accept the universality of the set of cardinal numbers, but we need to agree that there is a common core to the concepts that humans designate with numerals. Yet, while seemingly uncontroversial in the exact sciences, the cross-cultural universality of the set of whole integers has come under considerable assault over the past twenty years from relativistic anthropologists and sociologists.

Gary Urton's recent work on Quechua number and arithmetic, particularly concerning the *khipu* (*quipu*) knotted cords of the ancient Inka and modern Andean peoples, contains a number of stimulating ideas about the way in which number was used to express ideas of social harmony (Urton 1997). Urton asserts that Western concepts such as "odd/even" as we know them are not appropriate to the Quechua arithmetical experience. Rather, modern Quechua speakers employ an entirely different numerical dualism in which a set of integers such as {1,2,3 ... 100} is divided into *ch'ulla* {1,2,3...50} and *ch'ullantin* {51,52,53...100} (Urton 1997: 214-217). I do not agree with Urton, however, in his assertion that he has discovered among the Quechua an entirely new *ontology* of numbers, nor do I think that Quechua mathematical knowledge is irreconcilable with our own. No matter how creative and vibrant Urton's work shows Quechua arithmetic to be, Inka numbers can be treated in the same way as any others, and the Inka numerical notation system can be compared to others without any particular difficulty.

Jadran Mimica goes even further in his analysis of the numeral system of the Iqwaye of Papua New Guinea. Mimica found that the Iqwaye numeral words for 1, 20, and 400 (the last of which was only elicited from a single informant) are identical, each being represented by the term for "man" (Mimica 1988: 56). He then correlates the Iqwaye base-20 numeral system with the transfinite number series developed by Georg Cantor, in which x, x^2, x^3, etc., all have the same value, and concludes that the Iqwaye numeral system must in some way parallel that of transfinite numbers! I feel that this conclusion is insufficiently supported by ethnographic data and involves an erroneous leap of logic. It is one thing to refute the often ethnocentric assertion that members of a certain society are numerically incompetent, but quite another to imply that they are capable of mathematical thought well beyond most university-educated Westerners.

Relativist philosophers and sociologists of science have done much in the past quarter-century to attack the universalist conception of number. It is believed by some that the very foundations of mathematics can be assailed with a simple experiment. It is sometimes argued that 1+1 cannot equal 2 in any absolute manner, given, for instance, that if one were to take a cup of popcorn and add a cup of milk to it, the result would not be two of anything, but somewhat more than a cup of pulpy mush (Restivo 1992). Resisting the temptation to describe such casuistry as pulpy mush, I simply point out that addition is an arithmetical function that can only represent adding objects that are discrete and of a like nature. Counting is based on the perception of discontinuities or boundaries in reality (primarily but not exclusively visually), for which the discreteness of the objects in question is mandatory. The fact that we can *imagine* alternate ontologies of number does not imply that such ontologies are adopted in fact by members of any society, including our own. The evidence presented by Urton, Mimica, and others does not convince me that the number concepts of different societies are fundamentally incommensurable with our own. On the contrary, my own research suggests that the differences they have found are relatively inconsequential by comparison with the commonalities observed in all societies.

I acknowledge that, by treating all numerical notation systems purely as systems for representing number, I may not be able to do justice to the complex symbolism that underlies many of them. The arrival of the year 2000 was not simply another excuse to celebrate; rather, the nature of our numerical notation system and the "rolling-over" of the calendrical odometer on 2000/01/01 held great symbolic and even mystical significance for much of the world's population. The decision to underemphasize numerology in this study is partly a pragmatic decision based on space limitations. Yet it is also a statement that there is a *comparable core of features* underlying all lexical numeral systems and numerical notation systems, and that these similarities refute any relativistic arguments that would tend to deny the cross-cultural regularity of such systems. It is entirely likely that Quechua numerical symbolism is radically different from that of Han China, Ptolemaic Egypt, the Iqwaye, the modern West, or any other society. Yet these differences, while interesting, do not affect the essential validity of the cross-cultural comparisons I am undertaking.

Throughout this study, I will show that, while numbers and numerical notation are not necessarily used for functional or adaptive purposes, they are quite often used in mnemonic and arithmetical contexts that cannot be easily divorced from utilitarian concerns. It is a reasonable assumption that, while our hominid ancestors did not need and indeed did not use numerical notation, the ability to distinguish between two gazelles and three gazelles would have been cognitively important. No animal needs the concept of abstract number in order to survive. However, as the survival of early hominids was strongly predicated upon the ability to function in groups, it is reasonable to surmise that, along with more abstract understandings of space, time, and causality, the number concept would have developed early in human prehistory. It is highly probable that by the "cognitive revolution" of the Upper Paleolithic, *Homo sapiens sapiens* possessed languages including two or more numeral words and the ability to distinguish number even above the quantity expressible through language. Any strong form of cultural relativism cannot be upheld with respect to a domain of reality so crucial to our existence as number.

There are, of course, enormous philosophical problems underlying the ontology of number. In suggesting that the differences between numerical notation systems are not ontological ones, I do not mean to disparage in any way the work of the philosopher Gottlob Frege or his successors (Frege 1953 [1884]). Rather, I believe that the ontological problems of number are ultimately resolvable in a universal and psychological framework. In particular, where the philosophy of number is augmented by comparative psychological perspectives, insights can be reached regarding the way in which the human brain processes quantity that will be more fruitful than *a priori* assumptions about sensation and perception (Brainerd 1979, Hurford 1987). The fact that a wide variety of lexical numeral systems use addition, subtraction, and multiplication in similar ways to express complex lexical numerals suggests that, even if alternate ontologies of number are conceivable, the way in which all humans deal with integral quantities is fundamentally the same. It is on this basis – the common perception and understanding of number by human beings having the same basic brain structure and living in environments sharing features of quantity – that I can safely assert the universality of the concept of number as symbolized through numerical notation.

In the end, one might decide that the ontological and epistemological issues underlying the analysis of numerical notation systems are too difficult to enable comparison of any sort, much less the universal type of study I have undertaken. It could be claimed that each culture, and hence each numerical notation system, is a unique product of unique historical circumstances. If so, comparing Egyptian hieroglyphic numerals with Shang oracle-bone numerals and Inka quipu might be totally misleading. At best, even if there might be a core of features common to all numerical notation systems, I would be labelling oranges apples in order to compare them with other apples. At worst, if these systems are entirely different phenomena, I am trying to make apples out of abaci. To this argument, I can respond that the claim that all cultures are incommensurable is refuted by the relative ease of intercultural communication. The intercultural transmission of ideas relating to numerical notation systems is extremely frequent and poses a serious challenge to the relativist thesis. Prior to comparing phenomenon among multiple societies, we cannot assume either that the phenomenon is cross-culturally regular or that it is not. Having compared numerical notation systems on a worldwide basis, I regard the systems as being sufficiently similar to warrant their theoretical analysis as variations on a single theme.

Constraints, Universals, and Regularities

As I have established, the purpose of a universal comparative study is to determine, rather than to assume, the level of intercultural variability for a given phenomenon. Both the universalist position that human societies are highly regular and possess very many universal properties and the opposing relativist contention that societies are unique products of contingent historical circumstances are unsatisfactory, because they frequently presume rather than evaluate the degree of regularity that social phenomena display. In the study of numerical notation, the debate has been particularly muted, mainly because the universalist position has been adopted by historians of mathematics who, by and large, have been mathematicians rather than social scientists, while the relativist counterattack is fairly new and, at least so far, quite weak.[7] Throughout this study, I will demonstrate that, while numerical notation systems do display remarkable regularities and even universals, historical contingencies have also

---

[7] See Ascher 1991 and Urton 1997 for evidence that this trend may be changing.

played an important role in shaping the cultural history of numerical notation. I make no excuses, however, for my wholehearted acceptance of the psychic unity of humankind and the idea that cross-cultural regularities are of enormous theoretical importance for the discipline of anthropology.

Most anthropological theory is predicated on the existence of very strong constraints on the forms possible within human societies. While the statement that all human societies have kinship systems does not get us very far along in trying to understand culture, universals creep into every theoretical position within the social sciences. One could argue - and in fact it has been argued quite fervently by Clifford Geertz and others - that these universals are minimally true, but facile, irrelevant, and useless for understanding humanity (Geertz 1965, 1984). True understanding of humanity, it is argued, can only be accomplished through deep, long-term immersion in a single culture, in the course of which one acquires a true understanding of the tremendous power of culture to influence thoughts and behaviour. The insights to be drawn from cross-cultural comparison are trivial or 'fake', as they do not get to the heart of the human experience. There is much to be said in favour of deep immersion in a culture, be it through historical or ethnographic data. However, the denial of comparativism seems an overly negative position, given that those who criticize comparativism most harshly very often are those who have not undertaken it.[8] At its most extreme, this position requires that anthropology abandon theory entirely. In so doing, there seems a grave risk that it could lose its claim to be a unified scholarly discipline.

In his well-developed and cogent book on human universals, Donald Brown attempts to make anthropologists aware of the biological, psychological, and environmental conditions of human existence that result in universals (Brown 1991). By

---

[8] Sahlins (cf. Sahlins 1960 and 1976) is of course the primary exception to this generalization.

refuting many of the arguments of various brands of anthropological particularism (cultural-psychological, structural, and post-modern), he shows that, at least in broad outline, human societies share a common basic template. Without denying that this template permits tremendous variation in particulars, Brown argues that cultural anthropology is far too concerned with the study of variation and pays too little attention to the study of regularity.

Of course, numerical notation systems have been absent from most societies both historically and in the ethnographic present - although by now there is probably no society left on earth that has not had significant contact with the Western numeral system. In this sense, they are not cross-culturally universal as lexical numeral systems are. Nevertheless, this does not imply that there are no regularities to be discovered by comparing them, so long as we treat the numerical notation system rather than the society as the unit of comparison. First, there may be universals that apply to all numerical notation systems, for instance: "All numerical notation systems have a base of 10 or a multiple of 10". Such regularities are particularly notable in that they signal that there may be certain constraints on numerical notation systems that cannot be avoided. Second, there may be *implicational universals,* in which all numerical notation systems possessing a certain feature also possess some other feature, such as: "In all numerical notation systems with multiple bases, each of the lower bases will be divisible into the highest". These enable us to see certain features that are very common in a given sub-set of numerical notation systems, such as the five basic types of numerical notation system described above. Finally, there may be *diachronic universals,* which apply not to systems but to trajectories of systems. These regularities relate to similar patterns of historical development rather than synchronic structures. Any of these types of universals, if they have exceptions, become instead *statistical regularities.* While not true universals, statistical regularities are notable because they show constraints without implying determinism.[9]

One of the great contributions of anthropology should be to indicate the degree to which human societies are alike and the degree to which they differ. Nevertheless, I do not see the dichotomization of universalism and relativism as particularly useful. While acknowledging that some aspects of human existence are truly universal, and others are almost infinitely variable, most of the really intriguing questions facing modern anthropology require more nuanced perspectives concerning the sorts of constraints and inclinations that affect human societies. In the early 1900s, Alexander Goldenweiser developed his "principle of limited possibilities", which stated that for any social or cultural phenomenon, there is a limited number of possible forms that can be expressed in human societies (Goldenweiser 1913). Goldenweiser was particularly interested in the limitations imposed by human psychology on the expression of cultural traits. Although, given the inchoate nature of psychological theory at the time he was writing, he was unable to describe these mechanisms precisely, his insights are important in that they can be used to explain certain analogous developments in societies that are geographically and or chronologically distant from one another. More recently, Trigger (1991) has rejuvenated the idea of constraints in an attempt to revive the study of symbolic and mental regularities among archaeologists. In an attempt to reconcile the processual and post-processual camps in archaeology, he proposes that anthropologists should use the concept of constraint to describe all the limitations on human sociocultural variation – whether those constraints are biological, ecological, technological, informational, psychological, or historical. In so doing, he contends, we will be able to understand how it is that various factors can interact to produce the statistical regularities observed from

[9] For further discussions of types of universals, see Greenberg 1975 and Brown 1991: 39-51.

ethnographic and archaeological data without the need for determinism relying on one variable alone.

The insights of Goldenweiser and Trigger are very useful first steps in recognizing the importance of non-universal cultural regularities. The next step, of course, is to determine and describe the cognitive, functional, and environmental factors operating to produce such regularities. My primary concern with both the "limited possibilities" and the "constraint" approaches is that they are restricted by their formulation to considering the negative or restricting influence of various factors as being more important than the positive or imaginative effects. A simple reformulation of their insights notes that effects can restrict or limit options for cultural variability, but they may also in some cases positively incline societies towards a particular option. It seems important to differentiate between a very strong propensity in favour of some trait and a very strong constraint against all other possibilities. Constraints and inclinations can and do coexist, and the negative limitations of one variable must be weighed against the positive inclinations of another.

Turning specifically to numerical notation systems, I have used extensive textual and archaeological data to examine to what extent universals and regularities may be applicable. I follow Joseph Greenberg (1978), who in his analysis of significant regularities in lexical numeral systems developed a list of 54 universals and generalizations. Unlike much of his other work, Greenberg's study of numerals is universal and cognitive in orientation rather than phylogenetic. His work is synthetic, based on the detailed empirical work of earlier scholars, such as the German linguist Theodor Kluge, who spent decades compiling sets of numeral terms in languages throughout the world (Kluge 1937-42). While many of Greenberg's regularities have exceptions, are extremely complex[10], or are highly implicational, others reveal truly

[10] For instance: "37. If a numeral expression contains a complex constituent, then the numerical value of the complex constituent itself in isolation receives either simple lexical expression or is

universal and non-trivial features of every natural language; for instance, every numeral system contains a complete set of integers between one and some upper limit - each system is finite[11] and has no gaps (Greenberg 1978: 253-5). Out of a sample of thousands of natural languages, one never finds one in which "two" is expressed as "ten minus eight" or "twenty" as "one-fifth of one hundred".

While lexical numeral systems are found in every language, many human societies in the past have functioned quite well without numerical notation. It is possible to conceive of a world in which there are numerous regularities in lexical numerals, but in which numerical notation systems are highly specific and unique responses to local needs. However, we do not live in such a world. I will demonstrate that there is considerable uniformity among the world's numerical notation systems, and that they display many synchronic and diachronic regularities. I will return to this subject in considerable detail in Chapter 11, using the data assembled in the present study.

It is not enough to describe regularities; one must also attempt to explain the process or processes by which these regularities came to exist. It is my contention that the primary factors restricting certain types of numerical notation from common use and inclining humans towards other types can be derived from cognitive psychology. Because numerical notation is a means for representing number visually, the issue of how the brain perceives and conceptualizes quantity is very relevant to the evaluation of the kinds of numerical notation systems that can be developed, and more importantly, which kinds of system will be used in human societies.

The range of variation among the numerical notation systems is inherently far less than that which can be imagined by the human mind. To take only a very limited example, a numerical notation system can very easily be imagined that is just like the

---

expressed by the same function and in the same phonological shape, except for possible automatic phonological alternations, stress shifts, or overt expressions of coordination" (Greenberg 1978: 279-280).

[11] This is not true of numerical notation systems, some of which (like our own) are truly infinite.

standard Western system but instead of being a decimal system, having a base of any natural[12] number of 2 or higher. It is thus logically true that there are an infinite number of possible bases for numerical notation systems, and obviously, only a small number of those can exist among the finite number of human numerical notation systems. The fact that a large majority of numerical notation systems have a base-10 structure (and those that do not use multiples of 10) does not preclude the existence of binary and hexadecimal numerical notation for computing purposes. Similarly, while there are only five basic principles of numerical notation system found historically (as described above), it is easy to imagine other types that could have existed: a system where the size of a numeral-sign is relevant to its value, or where all composite numbers are expressed multiplicatively using prime number numeral-signs. A number of modern writers, abandoning traditional principles of numerical notation, have created new systems *ex nihilo* that rely on rather different principles than do the systems discussed in this study (Harris 1905; Pohl 1966; Dwornik 1980-81).

We are thus faced with a situation where the number and variety of conceivable numerical notation systems are far greater than what is observed or expressed in human societies. Why should this be so? One possibility is that something about modern Western society that has led scholars who have thought about the topic to imagine systems that no one from any other society could have imagined. Perhaps we are able to think of so many systems that are not historically attested because we have specialized technological needs - for example, for binary notation to aid in electronics. Perhaps our wide knowledge of other existing notations (past and present) grants modern thinkers a certain self-reflection regarding numerical notation. A cynic might even claim that the existence of professional academics like myself with nothing better to do with their time than to think of alternatives to our current system also could contribute to this capacity.

[12] Or even, as discussed in some aspects of number theory, having a fractional or negative base!

This argument is refuted in part by the large amount of tinkering with numerical notation that has taken place throughout human history. While there are few purely independent inventions, the adoption of both major and minor systemic alterations to existing systems is common in the history of numerical notation. It is thus quite erroneous to presume that only Western scholars are inventive enough to think of these logically possible but unattested systems. It remains possible that, because there is so much more modern scholarship than existed in any earlier period, fewer intellectual resources were spent on inventing numerical notation systems in the past.

A second possibility is that systems using unusual principles were invented, but that the effect of custom and tradition was so strong that the resulting systems were quickly rejected. If so, then it is possible that many systems existed in the past for which no evidence survives due to their rapid failure. Of course, we can only speculate about the existence of systems for which there is no evidence. Yet, in all cases where there is historical evidence of the invention of new systems followed by their rapid abandonment (e.g. 5th century Indian astrologers, 13th century Cistercian monks, 19th century Cherokee scholars), the systems in question fit into my typology.

A third possibility, one which I find most convincing at present, is that the human imagination is less constrained than are the functional requirements of a useful numerical notation system. While I do not deny that human imagination is itself constrained, I regard these constraints to be essentially unknowable, raising the philosophical problem of cognitive closure: that we cannot theorize about that which we cannot imagine. On the other hand, constraints on function relating to the actual uses of numerical notation - keeping in mind that "function" and "efficiency" are terms that can only be used in the context of a particular social or technical need - are knowable both through historical and ethnographic data and psychological experiment. By looking at the functions for which numerical notation is used, I will be able to show that certain possibilities are unworkable in practice.

Attempting to explain regularities in numeration from a constraint-based perspective allows us to speculate why certain numerical notation systems flourish while others do not. Is there some sort of "cognitive selection" by which individuals reject or filter out certain possibilities in favour of ones that are more conducive to perception and cognition with respect to certain societal functions? Are humans so bound by certain modes of thought that they are unable to break out of the mold, so to speak, and adopt numerical notation systems that do not conform to those previously in existence? Have we simply not been inventive enough (or around long enough) at present to conceive of an alternate workable system or principle? Such questions are, of course, tied up with the issue of whether one principle (such as the ciphered-positional of our own Western system) is 'superior' to others in any meaningful way, and they lead to the question of whether a truly new principle of numerical notation can emerge.

Cognition and Number

Cognitive psychology is the study of how the brain processes information. It includes, in particular, the study of sensation and perception, concept formation, attention, learning, and memory. Its methodologies are primarily experimental: because neuroscience cannot yet adequately observe the workings of the brain directly, cognitive psychologists study the brain by its observable outputs – the behaviour of humans under controlled conditions. Cognitive psychology regards information processing as crucial for human survival. Without the ability to form concepts, no sentient creature would be able to survive for very long. At the same time, however, it is readily acknowledged that these concepts are not perfect representations of reality. Firstly, the act of conceptualization and categorization requires that certain types of information be emphasized at the expense of other types. Secondly, there are errors in information processing that reflect the imperfect conceptual abilities of the human brain, which means that the concepts of the mind correspond only imperfectly with reality. In this regard, cognitive psychologists are in agreement with archaeologist Gordon Childe's argument that humans do not adapt to the world "as it really is", but rather to the world that they perceive as mediated through culture. Still, Childe insisted, human perceptions must correspond reasonably well with reality or else we would not survive (Childe 1956).

With respect to the study of numerical notation, we must concern ourselves with two intersecting topics within the discipline of cognitive psychology. Firstly, there is, quite obviously, the cognitive study of number - how the brain perceives visual quantity and uses this information to create the concept of number, and how in turn these concepts influence how people categorize external reality. Secondly, there is the important issue of cross-cultural cognition - how the culture in which individuals are raised affects the way in which they perceive and cognize information. Both topics have been extensively studied over the past twenty-five years.

One of the earliest works of cognitive psychology is in fact one of considerable relevance to the subject of the perception of quantity. Despite its age, Miller's seminal paper on the 'magic number 7 ± 2' remains an essential work for understanding how the brain processes number (Miller 1956). Miller's central argument is that in a number of related but distinct fields of human cognition, our capacity for processing information lies between five and nine 'units'. Two aspects of his research are particularly relevant to the study of number: subitizing (nearly-instantaneous perception of small quantities) and chunking (the organization of large amounts of quantitative information into smaller, more manageable units).

Firstly, using research conducted by Kaufman *et al.*, Miller discusses the process of subitizing, in which small quantities of figures or objects are perceived directly, while larger quantities must be encoded by counting, a more time-consuming process. This experiment involved showing groups of dots to subjects for 1/5 second, after which they would indicate how many were present; up to five or six dots, few errors were made (subjects were subitizing), while above that number subjects had to estimate and hence made more errors (Kaufman *et al.* 1949). In more recent studies, the limit of subitizing has been found to be somewhat lower than six, ranging around three or four for most experimental subjects under typical conditions (Mandler and Shebo 1982).

Closely related to subitizing is 'chunking', which distributes large quantities of objects among smaller groups, thereby enabling the brain to process the larger number as a certain number of the smaller sets rather than requiring each object to be cognized independently. North American telephone numbers of ten digits are divided into three "chunks" such as 414-595-2629 rather than written 4145952629, in part to distinguish the area code, local exchange, and individual phone line but also to facilitate memorization and recall. Chunking normally involves the division of a collection of objects into groups of three or four bits each, which, given that this is near the limit for human subitizing, speeds up the process of perception and accurate quantification by the brain. The perception of larger units as gestalts thus maximizes the brain's efficiency within the limits of its biological evolution.

Underlying these processes of quantification and enumeration is a single principle, *one-to-one correspondence*. One-to-one correspondence is the idea that numerical equivalence between two collections of things is established by pairing each object from the first collection with one from the second. As established by the developmental psychology of Jean Piaget, this is an ability acquired by human children around age four to six; before this point, children establish numerical equivalence by relying on perceptual cues such as spacing, and thus lack what Piaget calls 'conservation of number' (Piaget 1952). Adults use one-to-one correspondence when they hold up eight fingers to represent eight coconuts, when they put aside twenty-seven pebbles to count their flock of that many sheep, or when marking twelve lines on a sheet of paper to indicate the number of pints of beer consumed before staggering out of the local pub. Counting (as opposed to subitizing) cannot take place without one-to-one correspondence. One-to-one correspondence can be used in combination with chunking to increase the ease of representation and cognition. After my fifth pint, I might place a horizontal stroke through the four existing strokes to indicate a group of five; in so doing, my twelve pints would be rendered as two groups of five strokes followed by a group of two (probably rather erratic) strokes. By extension, numerical notation systems, particularly cumulative ones, rely on one-to-one correspondence.

Much of the debate on cognitive domains relating to mathematics and its origins takes place in the realm of comparative ethology, specifically studying number concepts in animals in order to create meaningful analogies with the abilities of human infants and adults (Fuson 1988; Gallistel 1990; Dehaene 1997; Butterworth 1999).[13] There is much skepticism about the ability of animals to count, which is certainly warranted given that the mathematical abilities of "Clever Hans" and other supposed animal calculators were shown to be the result of subconscious cues passed from human trainers to these purported prodigies (Fernald 1984). Following in the footsteps of Koehler's (1951) work on counting among birds, scholars are beginning to reach closer to home by studying primate numeracy, particularly the ability of great apes to subitize (Matsuzawa 1985; Boysen and Berntson 1989). Such controlled studies have shown that, although abstract mathematics is a strictly human province, many animals have certain abilities relating to the manipulation of quantity.

It is now generally agreed that various animal species are able to perceive quantity at least accurately enough to perform tasks involving small quantities, mostly up to three to five units. Furthermore, a general "accumulator" model has been developed by which many animals may perceive quantity, in which a counter in the brain records the accumulation of quantities up to a particular amount, though this

[13] These authors go into far more detail on the various research programmes undertaken to study animal and human infant perception of numerosity than is warranted herein, and review most of the relevant literature.

computation becomes increasingly fuzzy as quantities increase (Dehaene 1997: 23-31). It is not yet clear whether animal quantification is a primordial trait inherited from a now-extinct ancestral species, a convergent adaptation in many species to the requirements of similar physical environments, or a common cognitive response of animals having reached a certain threshold of brain complexity. Nevertheless, many independent experiments involving many different species have confirmed that something more than a "Clever Hans" phenomenon is being observed. The same appears to be true in the case of human infants; a large amount of research over the past fifteen years has suggested that even very young infants are able to distinguish small numerical quantities. Experimental research has shown that children as young as four months are able to distinguish one from two dolls, and can thus be said to have at least a rudimentary concept of quantity (Wynn 1992).

The ability to perceive numerical quantities (especially the lower natural numbers) is almost certainly a feature of evolutionary advantage to any species possessing it. While the ability to count and to formulate an abstract concept of number may be a universal human capacity, a product of millions of years of natural selection, the possession of numerical notation most certainly is not. Many societies (particularly small-scale ones) functioned very well for millennia without any need for a separate system for visually representing number. Furthermore, the variability among numerical notation systems cannot be explained fully by factors such as a universal human mathematical ability. Any human being (save perhaps those suffering from certain types of brain damage or other serious mental deficiencies) has the capacity to learn how to use numerical notation, but as a technology invented in particular historical contexts, its use is limited to those who have encountered it. Even so, this does not prevent us from considering the possible effects of human cognitive capacities on the types of numerical notation system that have been developed historically. It is very likely that the evolved capacity of primates to distinguish five from six bananas is significantly related to the human visual capacity to distinguish five from six strokes on a tally or knots on a cord. There are four biological characteristics of humans pertinent to the human development of the concept of number, which in turn is necessary for the development of numerical notation. These should be regarded as hypothetical, since they have not been confirmed by earlier research or my own, though I do not know of any evidence disconfirming any of them.

1. **Perception of discrete external objects**. The ability to distinguish foreground from background, to perceive the borders of external objects is necessary to the creation of the concept of "oneness". This capacity can be shown to exist in all animals.

2. **Perception and cognition of concrete quantity.** The ability to distinguish the quantity of sets of objects entails that a basic ability to quantify must exist. This does not necessarily imply a concept of abstract number, however. As discussed earlier, this capacity is present in many mammals and birds, and is generally restricted to low cardinal quantities.

3. **Possession of language.** Language, the fundamental human communication system, is tied to the development of a sequence of numeral words. The ability to identify numbers by linguistic symbols, as opposed to the pre-linguistic quantitative abilities possessed by infants and animals, permits communication about and conceptualization of number.

4. **Cognitive organization of quantities into a natural number line.** Humans, unlike any other species, must have the ability to think of number in terms of a set of quantities, each of which is separated from the one preceding it by one unit, up to some arbitrary limit. This ability creates a sequence of natural numbers.

While numerical notation systems are useful because they enable the human brain to perceive and cognize quantities efficiently, we must not assume that their structure and evolution can be derived *a priori* from the principles of cognitive psychology. Some neuropsychologists have attempted to examine the development of numerical notation from a cognitive perspective (Dehaene 1997; Butterworth 1999). Dehaene (1997: 115-117) uses a stage-based unilinear scheme to describe the development of numerical notation from its beginnings in simple one-to-one correspondence, through chunked groupings of notches and ciphered numerals to the ultimate stage of positional notation with a zero. However, we ought to be very suspicious of such schemes in the absence of significant historical documentation. While one-to-one correspondence may be more closely tied to human biology and cognition than other forms of expressing quantity, this does not mean that cognitive factors can fully explain the development of numerical notation. The contention that the history of technology can be understood as a sequence of ever-superior inventions "the better to fit the human mind and improve the usability of numbers" is still unproved and quite dubious (Dehaene 1997: 117). By assuming the course of the history of the development of numerical notation and then explaining this "non-history" cognitively, we fail to learn anything, except perhaps regarding the preconceptions of the researcher.

To establish the actual historical conditions under which numerical notation systems come into existence, we must consider additionally four sociocultural features that are necessary for any society to develop a numerical notation system. The sociocultural requirements for numerical notation systems are non-universal and derive from contingent historical circumstances. These factors are merely hypothetical, but since they are not universal, it may be possible for me to establish if in fact they are necessary conditions for the development of numerical notation using my universal cross-cultural methodology.

1. **Presence of organizing principles that structure the number line.** This feature involves the ability to structure the natural numbers in a manner most convenient to thought, and usually takes the form of a numerical base. There is no evidence to suggest that a numerical notation system has ever been developed in any of the world's

many languages in which there is no base, and in any event, the vast majority of lexical numeral systems do have a base.

2. **Presence of a non-structured tally-marking system based on one-to-one correspondence.** This feature is often claimed as the earliest stage of numerical notation through which all societies must pass. While this may be an overstatement, tallying is a very intuitive way to represent number visually. There is evidence for this form of representation as early as the Upper Paleolithic (Marshack 1972). It certainly has not disappeared and, of course, is widely used in the modern West.

3. **Social need for long-term recording of number (mnemonics).** The social need for a relatively permanent record of numbers seems essential to the development of numerical notation. If one of the main functions of numerical notation is to assist memory, the social need to remember numbers is probably necessary to its development. What we then must determine is exactly what types of society will have such a need, rather than assuming a single specific function in advance.

4. **Social need to transmit number outside a local cultural environment.** This is a loose requirement, but it is likely that numerical notation systems develop only where there is a need to communicate number outside a local community. While verbal numbers may suffice for local communication, the ability of numerical notation to communicate numbers across barriers of geography and language is an important feature that would make its development likely in such circumstances.

Because numerical notation is an invention of our species, it must be subject to the constraints imposed by our cognitive abilities. I thus deny that numerical notation can be studied without giving due consideration to the perception and cognition of visual quantity by the human brain. However, because it is a human invention deriving from specific historical contexts, the pattern of its historical development must be studied inductively before turning to universal cognitive approaches. Throughout this study, I will insist that while cognitive principles may constrain us from certain choices and recommend others to us, explanations derived solely from human cognition are unscientific and ahistorical. Moreover, a reliance on modern studies of cognition alone does not permit the examination of whether and how the cognition of number may vary cross-culturally and through time.

I believe that the adoption of numerical notation has important cognitive consequences for its users. These consequences, I suspect, are of a similar nature to Goody's (1977) suggestions regarding the consequences of literacy, though obviously restricted to the domain of number. Goody himself clearly believes this to be the case, as seen from his observations regarding the process of counting cowrie shells among the LoDagaa (1977: 12-13). The LoDagaa separate large groups of cowries into smaller groupings of five and twenty cowries to facilitate the counting of the larger group. While this is not numerical notation, as I have defined it, since it does not represent large numbers using new signs for a base and its exponents, it is an efficient way of counting a group of objects by dividing it into smaller equal groupings, and then counting the groupings. Goody notes that while LoDagaa boys were expert cowrie counters, they had little ability to multiply, a skill they had begun to acquire only recently with the introduction of formal schooling. He argues that, while both skills involve the manipulation of number, the former is concrete while the latter is more abstract and dependent on the introduction of literacy. The very existence of multiplication tables, a technique used by almost all Western children to learn to multiply, implies literacy and the use of numerical notation. Goody is careful not to overextend this distinction into an absolute dichotomy, but insists, quite rightly, that the formalization of numerical knowledge that accompanies written numeration has important consequences.

It is impossible in this study for me to compare the cognitive abilities of groups who lack numerical notation and those who possess it. This could only be done through the ethnographic study of a group before and after its members learned such a system. Here, I will be discussing only groups that possess numerical notation, and even then, I often do not have enough contextual information about how the numerals were used to draw cognitive conclusions. However, it may be possible for me to decide whether different types of numerical notation have different cognitive effects on their users. It is often assumed that cumulative systems such as the Roman numerals represent 'concreteness' in numeration because of their iconicity, while positional systems represent 'abstraction' because of their infinite extendability (Hallpike 1986: 121-122; Damerow 1996). Leaving aside cumulative-positional systems, whose existence is problematic for this dichotomy, such associations of numerical structure with cognitive ability are untested, and rely on the equally untested assumption that numerical notation develops from concreteness to abstraction over time. By examining the diachronic patterns in the evolution of numerical notation that actually occurred, I will be able to establish whether these patterns are unilinear and follow the proposed path. By comparing the structure of systems to the functions for which they were used, I hope to say something about the cognitive framework within which different groups understood number. Nevertheless, we should not assume that numerical notation alone can completely describe how different groups understood number, since it is only one part of a larger framework that includes mental calculation, lexical numerals, and other computational technologies.

Rather than assigning labels such as 'concrete' and 'abstract' to numerical notation systems, or identifying any other single factor on which the utility of a system should be judged, I will focus on a constellation of features of numerical notation systems that have cognitive consequences. This approach is similar to that adopted unsystematically by Nickerson (1988), who lists the relevant criteria as being: ease of interpretation, ease of writing, ease of learning, extensibility, compactness of notation, and ease of computation. Some of these criteria can be put more clearly into terms that can be evaluated: for instance, his 'ease of learning' (1988: 191) might better be decomposed into sign-count (the number of symbols used in a system) and iconicity (the degree to which numeral-signs resemble their values). Moreover, I do not think that 'ease of computation' should be considered as a factor, because the use of numerical notation systems for computation is relatively rare in pre-modern contexts. It is also desirable to know which of these factors might be more important than others, and this ranking may vary according to the functions for which a system is used. Having said this, a set of non-hierarchical criteria for evaluating systems from a cognitive perspective is a very valuable tool. Nickerson notes usefully, "If one accepts the idea that the Arabic system is in general the best way of representing numbers that has yet been developed, one need not believe that it is clearly superior with respect to all the design goals that one might establish for an ideal system. It may be, in fact, that simultaneous realization of all such goals is not possible." (Nickerson 1988: 198).

The notion that there is no ideal system but rather that each system is shaped by a set of goals which its users and inventors seek to attain, and which they can only achieve by compromising on other factors, is extremely important for analyzing the history of numerical notation. Without denying that there may be patterns of change among systems, it shifts the burden of proof to those who wish to maintain that numerical notation evolves in a unilinear sequence. When I return in Chapter 11 to the cognitive analysis of the systems I have studied, the importance of this perspective will become clear.

Numerals and Writing

The scholarly analysis of numerical notation often has been pursued by scholars interested in writing systems. Therefore, numerical notation systems are usually regarded as a subcategory of writing systems (Diringer 1949; Daniels and Bright 1996; Harris 1995). This is perfectly understandable, since most numerical notation systems are associated with one or more scripts, and conversely most scripts have some special form of numerical notation. The process of recovering instances of numerical notation archaeologically and interpreting them will thus inevitably involve epigraphers, paleographers, and other scholars of writing. However, the uncritical acceptance of a close connection between numerical notation and writing can lead to unfounded assumptions regarding their nature and the relations between them. I therefore think it pertinent to examine this connection more closely.

There are three basic ways that number is expressed by human beings: a set of verbal numeral words, the written expression of those words in phonetic scripts, and the signed expression of number through numerical notation systems. We can divide these three types into *auditory* systems (verbal lexical numerals) and *visual* ones (e.g. written lexical numerals and numerical notation). Alternately, we might distinguish *lexical* (verbal and written numerals) from *non-lexical* (numerical notation) means of expressing number. If the similarities between the two visual representations are more significant than the similarities between the two lexical representations, then the connection between numerical notation and writing is strong. However, several important differences between lexical and non-lexical representations of number demonstrate that this distinction is the more important one.

One point of contrast is that lexical numerals are a system for representing language, while numerical notation represents number through non-linguistic signs. However, the distinction between "writing" and "not-writing" is an issue of great debate among modern scholars, particularly in Mesoamerican (Marcus 1992) and Andean (Urton 1997, 1998) studies. The most restrictive approach holds that only systems that represent phonemes are scripts. Accordingly, the Maya glyph system is a "true" script, while the Aztec system is a semasiographic system that requires a great deal of context in order to be interpreted, and the Inka *quipu* notation is a numerical notation system with some small non-numerical component. A broader approach holds that phoneticism is not an essential feature of scripts, and classifies pictographic representational systems, numerical notation such as *quipu*, and even pictorial art under the rubric "writing". Gelb's classic definition of writing as "a system of intercommunication by means of conventional visible marks" (Gelb 1963: 253) would suggest that a numerical notation system *is* a script. However, I do not believe that Gelb considered numerical notation systems to be scripts in their own right. I am sympathetic to the argument that classifying societies as illiterate can be used to denigrate their members' intelligence and inventiveness, and that a broad definition of writing helps to counteract ethnocentric assumptions. Even so, I think there is enormous theoretical value in recognizing the distinction between phonetic and non-phonetic representations. I do not consider numerical notation to be "writing", and when I use the terms "writing" and "script", I am referring only to systems that represent phonemes. Nevertheless, my statement that scripts are phonetic and numerical notation is not is a definitional assumption, not an empirical observation of difference.

Because numerical notation is non-phonetic, it transcends language and can traverse linguistic boundaries. This does not imply, however, that numerical notation systems need no interpretation or translation. To an individual or group unfamiliar with our numerals, the specific meanings attached to each symbol are inscrutable without assistance - unless, of course, they are encountered in computations that permit the decipherment of their meanings from their context. Even systems that use one-to-one correspondence to represent numbers require that individuals know how the numerical base of the system is structured and what each symbol represents; a vertical line or bar represents one in many Old World civilizations but means five in Mesoamerica. However, once an individual learns a numerical notation system, he or she can communicate numbers with any other individual familiar with the system, regardless of their linguistic differences. Numerical notation is thus a fundamental aspect of international trade and the administration of large empires.

Another theoretical contrast concerns the contexts in which written lexical numerals and numerical notation are encountered. It would be an error to assume that all numerical notation is written. Numerical notation systems are not in any way limited to societies possessing scripts, nor do societies with scripts necessarily possess numerical notation systems. Unfortunately, while scholars such as Ifrah (1998) and Guitel (1975) mention the existence of tallies, knotted strings, and other such technologies, they are considered solely as peripheral (and often ancestral) phenomena to numerical notation proper. However, the *quipu* knot records of the Inka (ch. 10) and the tally-sticks of a wide variety of literate and non-literate civilizations lie within the scope of this study. One problem with studying such systems is that much numerical notation is notched on wood, drawn in sand, or knotted on ropes or strings, all of which are unlikely to survive archaeologically, while formally written numerical notation is often found on durable metal, stone, or clay. Far more numerical notation once existed in non-written contexts than has survived to the present. Any numerical representation that is visual and primarily non-phonetic is numerical notation, regardless of the context in which it was inscribed.

Moreover, just as numerical notation is not necessarily encountered in conjunction with writing, many scripts have no corresponding numerical notation system. For instance the Ogham script of Ireland, the Canaanite script, the early alphabets of Asia Minor such as Carian and Phrygian, and the indigenous scripts of the Philippines all lack numerical notation and always express numbers lexically. This shows that written lexical numerals and numerical notation are useful for different purposes. In societies that possess both scripts and numerical notation systems, there are strong norms prescribing the means of representing number depending on social context. For instance, throughout the Western world, lexical numerals are preferred in literary or religious contexts, while numerical notation is preferred in commercial transactions and accounting. In cases where both systems are found in a single text, there is often a functional division between the two; for instance, the text of the Bible is written using lexical numerals, but chapters and verses are numbered using numerical notation. In writing cheques, numerical notation predominates, but dollar amounts are written out in full to prevent forgery. Such contrasts in the functions of the two means of representation suggest that they should be treated separately.

Another distinction between lexical and graphic representations of number is that numerical notation systems and scripts exhibit very different patterns of geographical distribution and historical change. In part, this may be because scripts are at least partly phonetic, and their diffusion across space and time can be constrained by patterns of language use. In contrast, numerical notation is non-phonetic and trans-linguistic. This fact alone allows a numerical notation system to diffuse more readily than a script, which of necessity represents certain sounds and not others. Our own numerical notation system, Western numerals, derived initially from India and passed through the Arab world before reaching Europe, while our script is the Roman alphabet, of Greek and Phoenician ancestry. This historical differentiation is not uncommon. I will show throughout this study that the path of diffusion of numerical notation was in some cases radically different from that of the diffusion of scripts. Yet I do accept that in many cases there may be a connection between the indigenous development of writing and numerical notation. In several historically unrelated cases (Mesopotamia, Egypt, China, and Mesoamerica), the independent invention of numerical notation coincided closely with the development of a full-fledged script. The earliest proto-writing in of all these civilizations contains numeral-signs. Perhaps the need for numerical notation and a phonetic script tends to arise under similar circumstances (namely, during the formative phases of early civilizations, as in the four cases mentioned above). It could also be argued that the idea of numerical notation is one that, once developed, suggests to its users that other domains of activity might also be represented visually.

Yet another contrast arises when examining the structures by which written lexical numerals and numerical notation express number, which are quite different. The simple fact of being denoted visually is not as important as the different principles used in the two symbol systems. Lexical numerals (whether written or verbal) share a common structure that is very different from that of numerical notation. For instance, while the cumulative principle is very commonly employed in numerical notation, it is largely absent from lexical numeration. No known language expresses 'thirty' as 'ten ten ten'. In lexical numeral systems that have a base, multiplicative-additive structuring is overwhelmingly prevalent, whereas numerical notation systems are only rarely multiplicative-additive.

To take an extended example with which we are all familiar, let us compare Western numerical notation with North American English numeral words. Our numerical notation system is purely base-10 and ciphered-positional, and can be used to express any integer, since one can add zeroes to the right of a number *ad infinitum*. Our lexical numeral system, however, is decimal-millesimal, as our numeral words are structured using a mixed base of 10 and 1000 (*one million = 1000 x 1000; one billion = 1000 x 1000 x 1000*), and it is multiplicative-additive. The situation becomes even more complex if we include British English, which has a mixed base of 10 (ten tens = one hundred), 1000 (one thousand thousand = one million) and 1,000,000 (one million millions = one billion). Furthermore, while our lexical numeral system is *potentially* infinite, one needs to develop new words to express higher and higher values. The highest number in many English dictionaries is *decillion* (10^{33} in American English, 10^{60} in British English). There are also irregularities in our system: for instance, *eleven* and *twelve* do not follow the regular pattern for numbers between 13 and 19, and words like *dozen* and *score* add further complexity. Thus, while our lexical numerals and numerical notation seem complementary, closer examination shows them to be different in structure. If we were to look at other languages – even ones closely related to our own such as Danish – we would find that English is actually fairly regular in terms of its correspondence of lexical numerals and numerical notation (Menninger 1969: 65-66).

The comparison of scripts and numerical notation systems becomes more complex when dealing with scripts relying heavily on logography, such as Egyptian hieroglyphs or Chinese numeral characters. While many logograms in such scripts have a phonetic component, there is no requirement that *every* logogram must do so. In fact, logographic scripts rarely represent any phonetic element when writing numbers and rarely distinguish between lexical numerals and numerical notation. Thus, while in English, a given number can be expressed either using lexical numerals (*four hundred and fifty-one*) or in numerical notation (*451*), the same number has only a single expression in traditional Chinese characters (四百五十一). I recognize that a tricky definitional issue arises from this convergence of numerical and writing systems, and consider Chinese numerals to be both part of the Chinese script and a numerical notation system in their own right. Nevertheless, I insist on treating the numerals used in logographic scripts as numerical notation. They conform to the structural principles of numerical notation outlined earlier, and often differ from the structure of the verbal numeral words in the languages that the script represents. Furthermore, in many cases, their structure may change over time under the influence of other numerical notation systems, or they may be adopted by societies that use non-logographic scripts and that express lexical numerals in writing.

Numerical notation systems are thus semasiographic systems. Semasiographs depict ideas in a conventional and non-linguistic form; they range from the conventional signs for the suits used on playing cards to the indicators used to warn of biohazardous or radioactive materials. However, numerical notation is a very complex *system* of semasiographs: individual numeral-signs are essentially meaningless except in relation to the system of which they are part. Semasiographic systems are normally restricted to recording information about a specific domain; other such systems include ones for recording music, dance, or genealogy. No other semasiographic system has ever achieved the frequency of use of numerical notation systems, either in any specific society or on a worldwide basis. It is notable that number, of all domains that can be represented visually, should be so commonly represented semasiographically. Why this should be is a complex question. Partly, it may be that number is discrete and thus more easily conceptualized in visual terms - one dot for one sheep is a very basic analogy, as demonstrated by the presence of numerical concepts among various animal species and human infants. However, I suspect that the sociocultural *need* to represent number is largely responsible for its historical prevalence in comparison to other systems. Since my analysis is restricted to the domain of number, I cannot answer in any final way why this should be true.

The distinction between lexical and non-lexical representations of number is significant. To analyse numerical notation systems as adjunct components of scripts does not do them justice. Nevertheless, throughout this study I will sometimes refer to numeral-signs and numeral-phrases as being 'written'. When I do so, it is mere conventionality, and this usage does not indicate any specific relationship between numerals and scripts.

Diffusion and Invention as Evolutionary Processes

Numerical notation systems develop out of purposeful human efforts to perform certain tasks related to the visual representation of number. The central issue, as I see it, is not simply to understand the history of numerical notation as a sequence of historical events, but rather to explain the origin, transformation, transmission, and decline of systems. This does not mean that technical and functional aspects should be given priority over social and ideological factors; rather, social context and historical contingencies must be incorporated into adequate analyses of the histories of systems. It does require, however, that I distinguish *analogies* and *homologies*. Analogies are instances in which something similar develops independently or convergently two or more times, while homologies are similarities that derive from the descent of two or more cultural forms from a common ancestral society.

One rephrasing of the analogy/homology contrast results in the age-old dichotomy of independent invention and diffusion. Anthropologists who hold a materialist perspective on invention argue that humans are generally rational and creative. The most radical forms of such arguments derive from the 'cultural materialist' framework of Marvin Harris (Harris 1968). Cultural materialists usually explain social change in terms of adaptation and cultural evolution, and strongly prefer analogical explanations to homological ones. In such theories, inventiveness is only limited by certain constraints imposed by pre-existing technical or social forms; not even the most radical cultural materialist would argue that a mobile hunter-gatherer group could develop parliamentary democracy or repeating rifles. In the study of numeration, analogical claims include the assertion that the use of strokes for units should not be used as evidence of historical connection, but is an idea that comes naturally to the human mind (Ifrah 1998: 391).

On the other side of the debate are diffusionists[14], who assign priority to homologies and view cultural developments as the results of unique and contingent historical sequences. Perhaps the most prominent anthropological proponent of homological explanation is Driver (1966), who found that historical factors were much more important than functional ones in explaining kin-avoidance behaviour among North American Indians. Excessive reliance on homological explanation leads to hyperdiffusionism, which reached its greatest prominence in the 1920s among

[14] I group "migrationists", who explain homologies in terms of movement of people, together with diffusionists, who emphasize the movement of ideas from one society to another. In terms of the debate I am discussing, it is unimportant whether the movement of people accompanies the movement of ideas, because independent invention is seen as relatively unimportant regardless.

ethnologists such as Elliot Smith (1923) and Perry (1923, 1924), who contended that all higher culture derived from ancient Egypt. Hyperdiffusionism is relatively common among scholars of numeration, and is typified by claims such as Seidenberg's (1986) insistence that the Maya use of the concept of zero derived from the somewhat similar Babylonian concept.

Anthropologists generally agree that neither independent invention nor diffusion can be regarded as epiphenomenal in explanations of cross-cultural variability in human societies. Yet the longstanding scholarly effort to determine which is more important, or rather where on the continuum of independent invention and diffusion the reality of a situation may lie, has not been particularly fruitful (Steward 1929, 1955; Kroeber 1948; White 1962; Tolstoy 1972; Driver 1966; Jorgensen 1979; Maisels 1987; Burton *et al.* 1996). I suspect that the question is being formulated improperly and that it rests on a false dichotomy between analogy and homology. Diffusion and independent invention are two distinct processes by which innovations are introduced into societies, but they are complementary, not antithetical. I insist that both analogical and homological explanations must be incorporated into any model of the cultural evolution of numerical notation. Moreover, simply classifying an innovation as representing either diffusion or independent invention is a futile task. We need to answer specific questions about these processes. What sorts of inventions are developed, for what reasons, in what social contexts, and by whom? Which inventions are likely to be transmitted, how does this take place, and why are they adopted in some circumstances but not others?

On the surface, my suggestion that diffusion and invention must both be understood as evolutionary processes is similar to Harris' (1968) cultural materialist arguments on this subject. Harris contended that the idea that diffusion and independent invention are two different types of explanations for cultural features was a pernicious myth promulgated by Boasian anthropologists as part of their particularistic framework in order to deny the reality of evolutionary patterns. Instead, he proposed that independent invention was extraordinarily frequent, while diffusion was a sterile 'nonprinciple' that was "not only superfluous, but the very incarnation of antiscience" (Harris 1968: 378). Because both diffused and invented features must be accepted into the adopting society, he argued that nomothetic principles alone suffice to explain both analogies and homologies. Wherever an innovation takes root under conditions of culture contact, it must be an adaptive solution to a pre-existing problem in the recipient society. Thus, diffusion cannot be understood as a fundamentally different process from independent invention; both are part of larger causal sequences by which cultural phenomena are developed, altered, transmitted, accepted, rejected, or abandoned (1968: 377-78).

While I share with Harris the belief that nomothetic explanations are important, and agree that the dichotomy between diffusion and independent invention is not fruitful, I do not think that diffusion is a nonprinciple or that historical explanations are unimportant. Harris was reacting, quite understandably, against the strongly idiographic theoretical perspective among Boasian anthropologists. He was right in his contention that simply invoking diffusion as an explanation, without examining the processes by which innovations come to be accepted into a society, is no explanation at all. However, in practice, his rejection of diffusion led him (and most other cultural materialists) to ignore diachronic processes of change and cultural contact and to assume without warrant that cultural adaptation is a unitary process and that analogical explanations are the only ones worthy of scientific consideration.

Notwithstanding, there is both empirical and theoretical value in historical explanations. If we were only interested in understanding the adaptive reasons for innovations, diffusion's role would be minimized, but anthropology and archaeology are also historical disciplines, seeking to describe as well as to explain. The cultural materialist position assumes without proof that whether an invention is diffused or independently invented is essentially irrelevant and uninteresting, and thus each case of innovation, whatever its source, should be treated as an independent recurrence of cause and effect. This presumes that human beings are infinitely rational, that they immediately and fully recognize social problems, and that innovations will not vary according to how and by whom the solutions are introduced. However, if traits produced by homology are fundamentally different than those produced by analogy, these presumptions are refuted.

My theoretical framework is one that regards independent invention and the acceptance of diffused innovations as interrelated aspects of a generalized innovative process. Both independent invention and the acceptance of transmitted inventions require the rejection, on both the individual and societal level, of existing prejudices against change and, in overcoming this cultural inertia, perceiving the advantages of new ways of doing things. The idea that independent invention reflects human nature and functional adaptation, while diffusion spreads existing cultural traits adventitiously across space and time, is not substantiated by the evidence from numerical notation systems. However, this admission does not allow us to ignore cultural contact as a meaningful explanation for why cultural change occurs. One of the major advantages of my approach is that it permits me to treat homologies and analogies on the same basis, while not denying their differences, but rather highlighting the ways in which they may be alike or may differ. The more important question is not "Is this system's development a case of independent invention or of diffusion?" but "What roles do invention and diffusion play in explaining how this system came to exist and be adopted?"

'Diffusion' is often implicitly taken to represent a more or less benign transfer of information or technology from one group to another, followed by a period in which the recipient society evaluates whether there is a need for the innovation, followed by its acceptance or rejection. This is an extremely naïve view of processes of knowledge transmission under conditions of cultural contact, one that negates entirely the role of imperialism, peer-polity networks, and trans-societal institutions. For instance, many numerical notation systems developed in societies just at the time when they began to enter into long-distance trading relationships with societies that already possessed numerical notation. In these cases, cultural contact was the precipitating event behind the innovation, since the purpose of numerical notation was to facilitate this new social development. In such cases, diffusion does play an enormous causal role; without it, the change would not have occurred. Diffusion as the explanation for social change is even more important when we consider encapsulation, conquest, and other situations by which a stronger society is able to impose social institutions on a weaker one. However, such situations do not result inevitably in the direct adoption of the stronger society's system.

While historians, anthropologists, and archaeologists have abandoned the rather facile diffusionary theories and assumptions of the past (e.g. hyperdiffusionistic correlation of traits, the 'age-area' principle, and the use of migration as a causal explanation for culture change), the study of numerical notation remains riddled with the methodological assumptions of the past. Nevertheless, if we were to abandon diffusion entirely, we would be relinquishing a valuable theoretical distinction between analogy and homology. The existence of historical contingencies need not be fatal to the development of a cultural-evolutionary theory of numerical notation. In fact, in order to demonstrate empirically the cultural evolution of numerical notation, we must examine systems in a phylogenetic perspective that examines how systems change in a patterned way over time. We need to rid ourselves of the notion that historical explanations must be particularistic explanations, and determine empirically how analogical and homological processes account for specific developments. To do so, the conditions under which numerical notation systems are invented, transmitted, and adopted need to be compared rigorously. In so doing, I hope to produce a general theory of the cultural evolution of numerical notation systems that transcends the specific circumstances of any one case study.

In the perspective I adopt in this study, there are four basic questions to be answered regarding the development of each numerical notation system:

1. What antecedent(s) does the system have? As discussed above, all systems have social and technical prerequisites that are necessary conditions for their development: there is no truly antecedentless invention. The question I am asking here, however, is whether a given system is historically descended from one or more antecedent numerical notation systems. Numerical notation can be documented as having independently invented about half a dozen times, and these 'pristine' systems stand at the head of cultural phylogenies. Independent invention should not be the null hypothesis for any account of the origins of a system, but neither should it be restricted only to very ancient systems. Most systems have one antecedent only, while a few systems have two.

2. Does a system have an endogenous or an exogenous origin (i.e. was the stimulus to its development internal or external to the society in which it was invented)? The issue of whether or not each system's antecedent was located within the society in which it developed or was transmitted through intercultural contact is probably of limited significance for a theory of innovation. It is essential, however, for the historical perspective I am adopting, because it establishes the paths of transmission of systems within each phylogeny.

3. Does the new system supplant (in part or in whole) one or more older systems? I want to know what happens when a new system is introduced into a society that already uses numerical notation. Four outcomes are possible: a) the older system may replaced by the newly introduced one; b) the new system may begin to be used in conjunction with the older one, normally with some sort of division of labour between the two; c) neither the original system nor the newly introduced one may be accepted; rather, elements of the two may be commingled to create a third system; d) the new

system may be rejected entirely, while the older system is retained. All these outcomes are attested multiple times.

4. Does the new system use the graphic symbols and/or the structural principles of its antecedent(s)? I want to know in what specific ways the new system resembles its antecedent(s), either in the form of its numeral-signs or in its structure (base, interexponential and intraexponential principle(s), and additional signs). There will always be *some* resemblance, because I am not postulating historical connections between systems in the absence of such resemblances and related systems are often very similar. These resemblances, in conjunction with other historical evidence, will help delineate the nature of relations between the systems. Moreover, it may sometimes be possible to explain the differences between a system and its antecedent in terms of the functions for which they were used, the media upon which they were written, or pre-existing sign systems in the recipient society.

To answer these questions adequately, it is necessary that some criteria be adopted to distinguish endogenously invented systems from ones introduced from outside a society, and in the latter circumstance to identify as certainly as possible the specific ancestor-descendant relationship involved. Discerning historical relations among cultural phenomena can be extremely contentious, particularly when only archaeological data are available to demonstrate such connections. Should we, as Rowe (1966) suggests, restrict ourselves to diffusionary explanations only when there is abundant evidence of colonies, trading posts, or traded objects that independently confirm that two regions were in contact? Alternately, as proposed by Tolstoy (1972), is it sufficient to show that a particular combination of features is unlikely to have occurred independently more than once in order to demonstrate that cultural contact between the two regions must have occurred? This issue is not easily resolved, and may be unresolvable in the abstract because the ease of demonstrating the transmission of cultural features between regions depends on the nature of the phenomenon being studied.

Few inventors of numerical notation systems have ever provided detailed information about the systems that have influenced their invention. We know that the Cherokee numerals (ch. 10) developed by Sequoyah were invented on the model of Western numerals because their inventor told us so, but this is a rare instance. Thus, for most systems, I am forced by the limited surviving data to build a circumstantial case for their origins. In order to demonstrate cultural affiliations between numerical notation systems, I will use a set of criteria that involve both internal (structural and graphic) resemblances between systems as well as external (contextual and historical) considerations. The nature of numerical notation systems is such that they are always combinations of a number of traits, any of which may not be particularly unlikely to occur on its own but which become increasingly unlikely to have recurred independently in combination. Where there are a number of internal resemblances between two systems, it becomes increasingly likely that the earlier system is in some way ancestral to the later one. If the number and nature of these resemblances is large, cultural contact may be postulated in the absence of external confirmatory evidence. However, such arguments are inherently weak, because they require that we believe that only a single feature of a society (a numerical notation system) was transmitted from one region to another, while we would expect other signs of cultural contact. In no case, however, do I postulate a connection between two systems solely on the basis that they were used at approximately the same time and in a single region. There must always be some resemblance between postulated ancestor and descendant systems. In some cases, there may be two or more possible ancestors of a system, and there may not be sufficient data to decide which potential ancestor is more likely. This problem can be overcome with the accumulation of new data. In most cases other than the most clear-cut, it is essential to have both evidence from the systems themselves as well as historical and contextual evidence for the postulated event of transmission.

The main criteria I will use throughout this study in making arguments for historical connections between systems are as follows:

1. Use of the two systems at the same point in time.

This criterion is nearly unavoidable; there must always be some chronological overlap in the periods during which two systems are used for a hypothesis of cultural transmission to be sustained. It is remotely possible that a system that had gone extinct might be revived and modified by a later society (for instance, on the basis of old inscriptions), but this seems highly implausible and is hardly a sound basis for a hypothesis of cultural transmission. Alternately, it could be hypothesized that a system which is not attested to have survived long enough to be ancestral to another one did in fact survive; this is the basis of theories for the origins of the Etruscan numerals (ch. 4) out of the Mycenaean Linear B system. While such hypotheses cannot be dismissed immediately (especially if other factors suggest they could be true), in almost all instances, I require that there be some clear chronological overlap between a system and its postulated ancestor.

2. Similarity in structural features.

Because there are only three intraexponential principles (cumulative, ciphered, multiplicative), two interexponential principles (additive, positional), three common bases (10, 20, and 60), and two sub-bases (5, 10), no one aspect that is similar in two systems is sufficient to prove diffusion. However, when two systems are alike in all or most of these respects, cultural contact becomes a much more likely explanation for the resemblance. Many of the cultural families of systems that I will be discussing share a common structure; for instance, all the Italic systems (ch. 4) are cumulative-additive with a base of 10 and a sub-base of 5. This does not mean that all cumulative-additive quinary-decimal systems must be placed in that family – the Ryukyu *sho-chu-ma* numerals (ch. 8) clearly do not fit because they are used much later and have different numeral-signs. The use of structural features as evidence of contact also suffers from the weakness that, if many systems in a family are identical or similar, it is often impossible to choose between several equally likely candidate ancestors.

3. Similarity of forms and values of numeral-signs.

Because many graphic symbols are very complex, they are unlikely to be developed independently. If the forms of numeral-signs used in two systems are identical or very similar, *and* if those signs represent the same numerical values in the two systems, it is likely that cultural contact resulted in the invention of the later system based on the earlier one. The more signs that are shared between two systems, the more likely it is that there is a historical connection between them. However, when two systems use identical or similar signs for *different* numerical values, this is not evidence of such a connection. For instance,  and  represent 10 and 20 in the Kharosthi numerals (ch. 3) but mean 7 and 9 in the Brāhmī numerals (ch. 6). In this instance, even though the two systems were used in the same region at the same time (Mauryan India) and have two similar numeral-signs, the dissimilarity of the values of those signs does not allow this similarity to be used as proof of a historical connection. Caution must be exercised when invoking this criterion for very simple symbols: vertical and horizontal lines, dots, circles, crosses, and the like, because it is possible that they could recur multiple times. This is especially true in the case of the use of lines and dots with the value of one, since these signs may have been part of tally-systems before being used in numerical notation systems. Cases where signs are similar but not identical must also be treated with caution. There is no established body of theory for identifying relations among graphically similar signs; hence, such efforts usually proceed on an intuitive basis, often with the aid of historical data.

4. Known cultural contact between the regions where the two systems are used.

In general, where one feature can be shown to have been transmitted from one region to another, multiple features are likely to have been transmitted. Thus, where there is a known pattern of shared non-numerical features in two societies, or where there is substantial evidence of inter-regional trade, migration, or colonization, such evidence supports a postulated ancestor-descendant relationship between two numerical notation systems. Determining whether known cultural contact is sufficient to postulate the diffusion of a numerical notation system is always a tricky matter and involves an evaluation of various lines of evidence. For instance, one of the difficulties in postulating that the Brāhmī numerals (ch.6) are descended from the Egyptian demotic ones is that, despite structural and graphic resemblances between the two systems, Egypt is well down on the list of areas with which ancient India had contact.

This problem is made more complex by the concept of *stimulus diffusion*, first developed by Alfred Kroeber (Kroeber 1948: 368-370). Stimulus diffusion is a complex blend of inventive and diffusionary processes that results when awareness of an invention is transmitted from one society to another, but, because of some difficulty in transmission or acceptance, the actual invention does not take hold in the adopting society. However, because the general principle is seen as useful by the adopting society, some of its members, stimulated by the original idea, invent their own version of the invention. The most widely cited example of stimulus diffusion is the development of the Cherokee syllabary by Sequoyah in the 19th century, based on his rudimentary knowledge of the Western alphabet. Unfortunately, while several numerical notation systems resulted from stimulus diffusion (e.g. the abortive Cherokee numerals, never used in the syllabary), no body of theory exists to help identify stimulus diffusion using historical and archaeological data or to distinguish independent inventions, stimulus diffusion, and direct diffusion. It is sometimes tempting to postulate stimulus diffusion even when the basic fact of incomplete transmission cannot be established, especially when minimal evidence exists of direct cultural transmission as well as some resemblances between two systems. One hypothesis offered by Needham for the origins of Shang numerals (ch. 8) is through stimulus diffusion from Mesopotamia (Needham 1959: 149). However, in this study, I will use stimulus diffusion as an explanation only when it can be established that the form of cultural contact that occurred between two regions fits Kroeber's model.

5. Use of ancestor and descendant systems in similar contexts.

If two systems are used for similar purposes, on similar media, or among similar social groups in their respective societies, this can serve as further confirmatory evidence that the two systems are indeed related historically. This factor, while useful, is never sufficient on its own to demonstrate such a connection, but it may provide further support. For instance, the spread of the Greek alphabetic numerals into Armenia and Georgia (ch. 5), though poorly documented, is confirmed not only by the striking similarities in the systems but also by the systems' use in Bibles and other liturgical texts.

6. Geographic proximity between the regions where two systems were used.

All other factors being equal, a system is more likely to have been modelled off one that is used by neighbouring groups than off one used more distantly. This is a particularly dangerous criterion to invoke, especially where cultural contact between neighbouring regions is less than with regions that are more distant. There are many attested cases where two very different and unrelated systems are used in proximity to one another, and many others where related systems are used at considerable distances from one another. Geographical proximity is such a weak measure of the likelihood of transmission that I will only use it as a last resort, and never as the sole factor for hypothesizing transmission.

By establishing links between ancestor and descendant systems within the limits of the available data, an enormous database of information on numerical notation systems develops, and systems can be placed in phylogenies detailing the connections between them. These data can then be used to draw comparative conclusions using the data from this study in order to analyse the evolutionary patterns of change in numerical notation systems in a direct fashion. These explanations are analogical, because they describe independent recurrences of cause and effect. However, they are also explaining homological processes resulting from cultural contact and the transmission of knowledge among many societies. This is a paradox only if we continue to accept that these two concepts stand in opposition to one another.

Technology, Function, and Efficiency

Through the efforts of historians of science such as Thomas Kuhn (1962), the notion of the unfettered linear progress of science has been demolished within the social sciences. At best, most historians of science and technology now agree, scientific "progress" can relate only to the solution of problems operant within specific frameworks of knowledge. Even so, it is undeniable that the rate of technological change and the increasing complexity of scientific and technological achievements over the past two centuries, and in particular over the past fifty years, is remarkable. Even diehard cultural relativists, who believe truth to be ephemeral and contingent, are forced to agree that there has been exponential growth in scientific and technological fields. Individuals in the modern West have access to unparalleled technological resources and are able to harness enormous quantities of energy. The rapid growth of technology is a correlate of our highly complex industrial society; by comparison, small-scale societies are limited in the scale and scope (though not the ingenuity) of their technical inventions. Without the social need for steam engines to transport raw materials, finished goods and labour, steam power remains a child's plaything, as with the aeolipile invented by Hero of Alexandria in the first century AD. Conversely, without the antecedent invention of the steam engine, not only would the Industrial Revolution have been stopped in its tracks, but also later inventions such as hydroelectric power would not have been possible.

These causal chains of inventions and their accumulating complexity as they are entwined in webs of social, political, and technological systems, are what constitute scientific and technical progress in the modern sense of the word.

The primary difference between 19th and early-20th century theories of technological and scientific progress and more recent ones is that the burden of proof now quite rightly rests on those who wish to demonstrate that such evolutionary developments have occurred. Social scientists no longer presume that the Western form of any institution, technology, or knowledge system is the pinnacle of human achievement. It is recognized that technological developments are related to the demands of particular social and technical environments. We no longer regard the development of agriculture, for instance, as the inevitable discovery of its many advantages over hunting and foraging, but acknowledge that it has borne a heavy cost for many individuals (Sahlins 1972; Boserup 1965). Without denying that Western technology has many advantages, and that such technologies could not have been developed in pre-industrial societies, we are wary of progressivist schemes that are constructed *a priori* to put all Western advances at the top of the ladder. Even where directional trends are evident, it is unacceptable either to impute moral superiority on the basis of more complex technology or to presume that no further developments will ever occur.

In the study of numerical notation systems, however, this shift in our conception of scientific and technological progress has not yet taken hold. Here is a small sampling of laudatory statements regarding our own (Western or Hindu-Arabic) notation found in the recent literature[15]:

---

[15] To spare the reader and to demonstrate the continuing existence of a historiographic problem, I have chosen only examples from the past twenty-five years, thereby avoiding the even more egregiously ethnocentric statements of some scholars from earlier decades.

> Sa perfection va bien au-delà de la civilisation indienne puisqu'aucune autre numération de Type III n'a jamais été en mesure de l'égaler (Guitel 1975: 758).

> When its advantages became apparent to the scholars and reckoners of civilizations in contact with India, they gradually abandoned the imperfect systems transmitted to them by their ancestors (Ifrah 1985: 459).

> If the evolution of written numeration converges, it is mainly because place-value coding is the best available notation. So many of its characteristics can be praised: its compactness, the few symbols it requires, the ease with which it can be learned, the speed with which it can be read or written, the simplicity of the calculation algorithms it supports. All justify its universal adoption. Indeed, it is hard to see what new invention could ever improve on it (Dehaene 1997: 101).

> Our positional number-system is perfect and complete, because it is as economical in symbols as can be and can represent any number, however large. Also, as we have seen, it is the most efficacious in that it allows everyone to do arithmetic ... In short, the invention of our current number-system is the final stage in the development of numerical notation: once it was achieved, no further discoveries remained to be made in this domain (Ifrah 1998: 592).

Historians of numerical notation overwhelmingly accept without proof that the Western numerical notation system is the most efficient ever developed. Moreover, it is quite possibly not only the "best" in existence, but also "perfect" – the best system that could ever be conceived. Its adoption by the vast majority of human societies today is perceived as a natural and inevitable consequence of this superiority, unmediated (or perhaps only minimally mediated) by social, political, and economic factors. Other, more cumbersome systems are to be evaluated in relation to the Western system. In particular, numerical notation is seen as a tool for performing arithmetical calculations and, inevitably from that fact, for developing higher mathematics. Mathematics is seen as a natural by-product of the rationalization of numerical notation over time. In addition, since so many modern conveniences and wonders are predicated on the existence of mathematics, Western numerical notation is in some sense a partial cause for these evolutionary developments. The corollary of this proposition, often left unstated, is that those societies that did not develop "our" way of numbering did not develop higher mathematics and did not evolve because of this fact.

I ought to make it clear at this point that I *do* consider the system of numerical notation developed in India in the seventh century AD and transmitted by Arab scholars to Western Europe to be a very remarkable invention. It does possess the properties of brevity, unambiguity, and ease of learning that make it conducive to the practice of written arithmetic and mathematics. It may very well be the most efficient system for representing and computing numbers. I also reject the idea that we cannot say anything at all about the efficiency of numerical notation systems. How well numerical notation systems represent number is one of many factors causing the development of new systems, their acceptance after being transmitted, their modification over time, and their eventual abandonment. Because of this process, a system that is inferior for a given purpose often will be gradually abandoned and replaced by one more suited for that purpose. This pattern of sociocultural change over long periods can meaningfully be called evolutionary.

The primary difficulty with the assumption of the evolutionary progress of numerical notation is *not* the notion of evolutionary progress itself. It is the idea that the efficiency of numerical notation systems cannot be evaluated in the abstract, but only in the context of the purposes for which a given system was developed and used. In particular, it is often assumed that the function of numerical notation is to perform written computations, either basic arithmetic or higher mathematics. Ifrah, whose work is the most popular and influential study of the history of numerical notation, makes this point abundantly clear:

> To see why place-value systems are superior to all others, we can begin by considering the Greek alphabetic numeration. It has very short notations for the commonly used numbers: no more than four signs are needed for any number below 10,000. But that is not the main criterion for judging a written numeration. What matters most is the ease with which it lends itself to arithmetical operations (Ifrah 1985: 431).

If, indeed, numerical notation systems have been developed largely in order to perform written computations, this would indeed be a fair basis for comparison. Yet I will show in this study that this view is entirely erroneous. The number of numerical notation systems used for computation is remarkably small. While numerical notation is probably a necessary condition for the development of mathematics, it would be Whiggish to argue from this that its purpose was to facilitate the development of mathematics. The efficiency of any technology can only be evaluated in terms of the purposes for which it was developed and/or used. There is thus no eternal abstract standard of efficiency for any technology. It smacks of teleology to argue that Western numerical notation is wonderful because it enabled modern mathematics to develop, when in fact the development of our numerals apparently had very little to do with mathematical computation and very much to do with writing dates on inscriptions in southern India and southeast Asia in the 7th century AD.

I will show throughout this study that the primary function of numerical notation is always the simple visual representation of numbers, apart from any considerations of efficiency for calculation. Most numerical notation systems were never used for arithmetic or mathematics, but only for representation, such as recording numbers unrelated to computation (as with most dated inscriptions, the most prevalent source of examples of numerical notation) or for writing the results of computations performed in the head, on the fingers, or with an abacus (as in most commercial computations). Without denying the enormous rise in the need for arithmetic in the modern West, computation remains a secondary function of numerical notation. On a not-so-crisp Canadian $5 bill, numerical notation is used to indicate a monetary value (5), the date the bill was designed (1986), a serial number with some letters to render it unique (GPA6537377), and the number 64 pencilled in one corner (no doubt to record the number of $5 bills received at some event). None of these numeral-phrases was actually ever used to *perform* arithmetical computation.[16] Numbers are used to denote far more often than they are to reckon, even in our highly numerical (though not necessarily numerate) society. This was doubly true in pre-industrial contexts.

In defining a numerical notation system as a visual and primarily non-phonetic structured system for representing numbers, I am explicitly making a functional statement. At its weakest, this statement implies that, whatever else numerical notation may mean in a particular society or to particular individuals, it must always express number as one of its functions. I would go further in stating that the primary function of numerical notation is always the expression of numbers. I do not regard this functional bias as a serious weakness. I am not saying that a numerical notation system must be fully integrated with other systems in a society, or that it must be perfectly adapted to serve the society's needs. Furthermore, I explicitly refute the notion that the purpose for which a technological innovation, such as a numerical notation system, was developed must remain that for which it is used.

Through this study, I will show the conditions that lead to the necessity or the desire for a numerical notation system, and discuss how these conditions may be similar or different from the actual contexts in which they are used. This is a troubling problem; some very complex societies (e.g. the civilizations of Teotihuacán and the Yoruba) appear to have done without any numerical notation system, or at best possessed an extremely limited one. Just as the old anthropological belief that all "true" civilizations must be literate has given way in recent decades to a more contingent view of the development of literacy, there is no simple correlation between a high degree of social complexity and the

---

[16] One might protest that the numeral on the bill is used in doing arithmetical computations such as providing change for purchases. To refute this, one need only go into a bank and ask for $100 in five-dollar bills, and see whether the cashier looks at the number on each bill, or whether in fact he/she merely counts out twenty bills while doing the arithmetic in his/her head. The numeral on the bill denotes its value, but is not used in calculation.

presence of numerical notation. Still, there is something to the statement that numerical notation systems only arise in the context of complex societies or in smaller-scale societies encapsulated, conquered, or in intimate contact with complex societies. Civilization may not require numerical notation, but perhaps numerical notation requires civilization.

Furthermore, the central function of numerical notation - expressing number visually - is general enough that it can be stimulated by a variety of social or political needs. While the need to enable financial transactions - by making a transaction possible over long distances, enabling the calculation of a monetary amount, or recording results to facilitate accurate bookkeeping - is the most obvious of such needs, it is not the only one. For instance, there is reason to believe that the main impetus behind the origin of the Mesoamerican numerical notation systems was astronomical and calendrical. The Indian ciphered-positional system, the much-vaunted ancestor of our own, appears to have been developed in order to reduce the number of different signs required to represent dates on inscriptions. One of the weaknesses of Denise Schmandt-Besserat's (1984, 1992) notable analysis of Sumerian tokens and their role in the origin of both numerical notation systems and scripts is that, although it is not cross-cultural, the accounting/commercial function she posits for tokens and subsequent account-records is incorrectly regarded as the universal cause for numerical notation. The scheme she proposes is unilinear and universal, without considering evidence from other societies that would suggest a more complex and contingent multilinear evolution of numerical notation.

If we do wish to compare the efficiency of various numerical notation systems, we must find systems that served a common purpose, and then evaluate them in terms of how well they served that purpose. When dealing with archaeological and historical data, the imputation of purpose is a complex question, but some general statements can perhaps be made. For instance, the one common purpose for which all numerical notation systems have been used is simple representation of number. Because this standard can be universally applied, it is a very good criterion for comparing different systems. One might further argue that a system that represents a number using few number-signs is more efficient than one that requires many signs. One could then argue that the Roman numerals are not as efficient for representing number as Western numerals are because 1492 is much shorter than MCCCCLXXXXII (or MCDXCII). In reality, the situation is more complex - MMI is shorter than 2001, for instance. Even if a person evaluated the length of all numeral-phrases in both systems in an attempt to find mean numeral-phrase length, one is faced with two problems. Some systems, such as our own, are potentially infinite in length, rendering the concept of "mean numeral-phrase length" meaningless. Moreover, even if some arbitrary cut-off point were assigned (say, all numbers less than 1000), different numbers are encountered more frequently than others in writing. It just so happens that, while Western numerals require fewer numeral-signs to represent small numbers and many other miscellaneous large numbers, the structure of the Roman system is such that round numbers (those that are exponents of the base or multiples of those exponents) are often represented more concisely than in Western numerals. In Chapter 11, I will attempt to resolve this issue at least partly so that systems' conciseness can be compared.

In the final analysis, it is clear that regardless of the frequency of various natural numbers in Roman and Western society, the Roman numerals are more concise for only a small fraction of all natural numbers[17]. Is the Western system more efficient for representing numbers than the Roman? This would be true only if we equate "efficiency" and "conciseness". There are many other criteria that could be used: ease of reading numbers, ease with which the system can be learned, whether or not the system can be

---

[17] Of all natural numbers up to 1000, the classical Roman system is more concise than Western numerals for only 15: 10 (X), 50 (L), 100 (C), 101 (CI), 105 (CV), 110 (CX), 150 (CL), 200 (CC), 500 (D), 501 (DI), 505 (DV), 510 (DX), 550 (DL), 600 (DC) and 1000 (M), although it is equally concise for many more (XI, XV, XX, etc.). If we were to consider the medieval/modern Roman system, which uses the subtractive principle, its conciseness would of course be significantly greater than the classical numerals.

infinitely extended, and, of course, the ease with which computations can be performed.[18] Even where there are definite answers to these efficiency-related questions with respect to specific numerical notation systems, this does not mean that individuals adopting or testing out a new system will immediately perceive the advantages and disadvantages of that system. A familiar but in some respects inefficient system, so long as it is not entirely unworkable, may be retained, despite the so-called "obvious" superiority of some other system. There may be a steep learning curve preventing the easy adoption of the alternative system, or there may be cultural or political reasons for retaining one's present means of representing number. There is also the problem that numerical notation, as a system for communicating information to others, requires not only that specific individuals adopt it, as would be the case with, a more efficient plough or a better mousetrap, but that an entire social group must learn and accept the new system before its usefulness will be evident. In this study, I will discuss many specific instances where cultural lag, a learning curve, or similar factors restricted the adoption of a so-called "superior" system, including our own.

The issue of efficiency of representation has been addressed in a rigorous if ultimately tendentious manner by Jiajie Zhang and Donald A. Norman in an important paper on the cognitive aspects of numerical notation (Zhang and Norman 1995). Zhang and Norman are not setting out to produce a typology of numerical notation systems for its own sake, but to examine the way in which specific systems visually represent (or fail to represent) information about the number(s) being expressed. They correctly identify the three general means by which numerical notation systems are structured: the **shape** of specific numeral-signs; the **quantity** of any particular numeral-sign within a numeral-phrase; and the **position** of numeral-signs within a numeral-phrase. They examine the way in which these features are combined in specific systems with the aim of undertaking a dimensional analysis of such systems. Systems are analysed with respect

[18] See Nickerson 1988: 189-197 for an in-depth discussion of these criteria.

to how many of the above criteria are used in structuring the base and power dimensions of each system, classifying them as 1x1D systems (those using two dimensions) or (1x1)x1D systems (those using three dimensions), as shown in Table 1.3.

Table 1.3: Zhang and Norman's Classification of Numerical Notation Systems: Examples

System	Type	Base	Sub-base	Power	Sample	GG	SC
Roman	(1x1) x1D	Q	S	S	678=DCLXXViii	IA	Cum-Add
Greek alphabetic	1x1D	S	N/A	S	678= $\overline{\chi}\overline{o}\overline{\eta}$	IB	Ciph-Add
Chinese	1x1D	S	N/A	S	678=六百七十八	IIB	Mult-Add
Babylonian	(1x1) x1D	Q	S	P	678= 𒌋𒁹 𒌋𒐁	IIIA	Cum-Pos
Arabic[19]	1x1D	S	N/A	P	678	IIIB	Ciph-Pos

Legend : Q = quantity, S = shape, P = position; GG - Guitel's classification (*Histoire comparée*); SC - classification presented here

Comparing Zhang and Norman's classification to my own, we can immediately see that the base dimension corresponds with intraexponential structuring, and the power dimension with interexponential structuring. In the base dimension, all cumulative systems use quantity, while in the power dimension all additive systems use shape and all positional systems, unsurprisingly, use position. One difference between our classifications is that Zhang and Norman do not distinguish between ciphered-additive and multiplicative-additive systems, in effect combining the two numeral-signs of each base in a multiplicative system into a single "shape-unit". Furthermore, in considering sub-bases separately, they add an additional dimension into their typology. While I do not deny that sub-bases are important structurally, I find it more useful to consider them separately from the main issues of intra- and interexponential organization.

[19] Zhang and Norman use the conventional term "Arabic" for our numerals where I use "Western"; I follow their usage where appropriate in discussing their analysis.

A further point in favour of Zhang and Norman's approach is that it attempts to correlate the principles used to structure numerical notation systems with principles that can be experimentally demonstrated to be operant within the human brain in structuring quantitative information. Their analysis compares those aspects of numerosity that are represented externally (through numerical notation) and those represented internally (in the mind). Numerical notation is a tool to enable *distributed representation*: some aspects of a task that would otherwise have to be represented internally, requiring cognitive resources, can be represented externally in order to make the task easier (Zhang and Norman 1995: 279-80). This approach requires that we consider both the cognitive capacities of the brain and the visual representation of number through numerical notation as a single system. In so doing, Zhang and Norman usefully recognize that numerical notation is a technology whose constraints are dependent on properties of human cognition, and that it can be studied using insights from the cognitive sciences.

While I am in general agreement with Zhang and Norman on taxonomic and methodological issues, I vehemently disagree with their conclusion that there is something special or unique about our own system. Their ingenious dimensional analysis represents very well how humans process visual numerical information, but does not demonstrate anything about the evolution of numerical notation because it does not consider the historical contexts within which systems were used. Because their paper begins with the claim that, "We all know that Arabic numerals are more efficient than Roman and many other types of numerals for calculation", their endeavour seems to affirm a long-held prejudice about numerical notation systems rather than to present a neutral account of the principles underlying such systems (Zhang and Norman 1995: 271). Other systems have not been tested against our own under experimental conditions (presuming that one could do so). Furthermore, the advantages they perceive for our system are only advantages vis-à-vis a set of untenable assumptions regarding the mathematical and arithmetical function of numerical notation. They conclude that the uniqueness of the Arabic system lies in the fact that for an arithmetical operation such as multiplication, far more of the information to be processed is represented externally (through numerical notation) than in any other system (Zhang and Norman 1995: 287-8).

Zhang and Norman wrongly assume that the criterion on which numerical notation systems are to be evaluated is their ability to perform arithmetical computations. Perhaps they are correct that the Western numerals are more advantageous than any other system for doing computations. Even if this were so, and I think that it would have to be resolved through actual practice and use of the systems rather than by relying on abstract principles, it would not be a fair question, because most other systems were never designed or used for such a purpose. The situation is analogous to denigrating screwdrivers for being inefficient hammers. The fact that one can use a screwdriver handle to drive in nails does not justify that comparison, just as the fact that one might use hieroglyphic numerals to multiply does not justify comparing them to systems such as the Western numerals. To add insult to injury, even though Zhang and Norman recognize that calculation technologies such as the abacus are demonstrably better than numerical notation for doing arithmetic, they suggest that part of Western numerals' superiority is that they are used for both calculation and representation, while other societies employed two separate systems (Zhang and Norman 1995: 293). They thus blame the carpenter for using both a hammer and a screwdriver where just the screwdriver would do. These arguments are little more than elaborate rationalizations for a historical fact (the near-universality of Western numerals) that eludes such a simple explanation.

* * * *

Having outlined the major theoretical issues pertaining to my study, I turn in the following chapters to the body of data itself. Throughout my presentation of these systems, I endeavour to highlight the ways in which the general theoretical principles discussed above may be confirmed or revised on the basis of empirical data. I have organized these data according to historical phylogenies or sequences of genetically related systems, so that the reader may follow the development of numerical notation throughout history. In each phylogeny, I have endeavoured to present the earliest systems first, leading forward to systems developed more recently. The first five phylogenies are probably related to one another historically, so I treat them together, but no other principle has been used in the ordering of chapters. The eight major phylogenies, each of which merits a full chapter, are as follows:

a) Chapter 2: Hieroglyphic family - systems historically descended from the Egyptian hieroglyphic numerals;

b) Chapter 3: Levantine family - systems used in the Levant, descended from the Aramaic and Phoenician numerals;

c) Chapter 4: Italic family - systems used in the circum-Mediterranean region, descended from the Etruscan numerals;

d) Chapter 5: Alphabetic family - systems whose signs are mainly phonetic script-signs, descended from the Greek alphabetic numerals;

e) Chapter 6: South Asian family - systems whose historical origins are in the Indian subcontinent and are descended from the Brāhmī numerals;

f) Chapter 7: Mesopotamian family - systems used in Mesopotamia, descended from the proto-cuneiform numerals;

g) Chapter 8: East Asian family - systems descended from the Shang oracle-bone numerals;

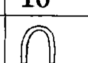
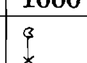
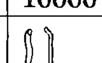
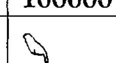
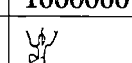
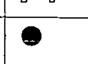
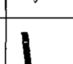
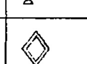
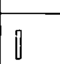
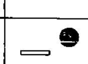
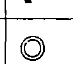
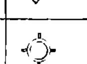
h) Chapter 9: Mesoamerican family - systems descended from the Mesoamerican bar and dot numerals.

Chapter 10 is devoted to miscellaneous systems and cultural isolates that do not fit into any of the above phylogenies, such as the systems of South America, West Africa, and the Harappan civilization. Chapter 11 analyses synchronic and diachronic regularities among numerical notation systems in a structural and cognitive framework, while Chapter 12 tempers these findings with important considerations relating to social context.

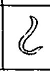
Chapter 2: Hieroglyphic Systems

The hieroglyphic family of numerical notation systems is one of the oldest and longest-lasting phylogenies in this study, with the Egyptian hieroglyphic numerals being used as early as 3250 BC. However, the cultural history of the numerical notation systems of the Mediterranean region has not been treated systematically. Thus, they have not previously been identified as members of a common family. I contend that a recognizable phylogeny of numerical notation systems was used in conjunction with a group of related "hieroglyphic" scripts and their descendants. Among these, I include the Egyptian hieroglyphic system, obviously, but also the Hittite hieroglyphic, Cretan hieroglyphic, Minoan Linear A, Mycenean Linear B, and Cypriote numerals. In addition, I include in this section the Egyptian hieratic and demotic systems, which are cursive reductions of the Egyptian hieroglyphic numerals, even though they are structurally closer to the alphabetic family of systems (ch. 5), to which they are almost certainly ancestral. In naming this family "hieroglyphic", I do not mean to imply that the scripts corresponding to these systems share a common structure, nor am I implying anything in particular about the nature of the numerals. Instead, I use the term simply because the earliest scripts associated with this family – in particular the Egyptian – are all known as "hieroglyphic" scripts and have a strong pictographic component. The systems of this family are summarized in Table 2.1.

Table 2.1: Hieroglyphic numerical notation systems[1]

System	1	10	100	1000	10000	100000	1000000
Egyptian hieroglyphic							
Cretan hieroglyphic							
Minoan Linear A							

[1] The hieratic and demotic systems are too complex to be included on this chart; consult their individual entries for their numeral-signs.

Mycenean Linear B	𐄇	𐄐	𐄙	𐄢	𐄫		
Cypriote syllabic	𐄇	𐄐					
Hittite hieroglyphic							

Of these systems, the Egyptian hieroglyphic has been most extensively discussed, though it has been repeatedly and profoundly misinterpreted in most histories of mathematics. The others are less well known, particularly the Cypriote system, which is neglected by Menninger (1969), Guitel (1975), and Ifrah (1985, 1998). Guitel discusses only the Egyptian hieroglyphs among all the above systems, which is consistent with her failure to consider systems that do not transform structurally over time or express some new principle. From a historical perspective, however, this family is extremely interesting, as it helps explain the spread of numerical notation systems throughout the Mediterranean region. The hieroglyphic family of numerical notation systems is directly ancestral to the Levantine (ch. 3), Italic (ch. 4), and Alphabetic families, but its systems differ sharply from those of its descendants.

Egyptian Hieroglyphic

The hieroglyphic script is the best-known ancient Egyptian script. It was used between about 3250 BC and 400 AD, making it the longest surviving of all scripts (Loprieno 1995). However, its use was restricted geographically to the Nile Valley and nearby areas under Egyptian control. While it remains possible that the hieroglyphic script arose because of stimulus diffusion from Mesopotamia, there is no solid foundation for such an assertion, since the scripts in these two areas emerged essentially simultaneously, with the Egyptian possibly slightly earlier. Hieroglyphic inscriptions are written from top to bottom, left to right, or right to left, with the last of these three options being the most common (Ritner 1996: 80). The script is mixed in principle, with both phonograms (consisting of one, two, or three consonants) and logograms indicating words non-phonetically (Ritner 1996: 74). The later hieratic and demotic scripts used to write the ancient Egyptian language, as well as the Meroitic hieroglyphic script, are directly derived from the hieroglyphic, while the early scripts of the Levant and the Aegean are probably its less direct descendants.

Numbers are very rarely expressed through phonetic numeral words in Egyptian, making it difficult to determine the structure of the lexical numerals, though evidence from some Old Kingdom Pyramid Texts and later Coptic writings has enabled linguists to establish the purely decimal structure of the numeral words (Loprieno 1995: 71). Most hieroglyphic inscriptions express numbers using seven ideographic numeral-signs representing the exponents of 10 between 10^0 (1) and 10^6 (1,000,000). These signs are shown in Table 2.2.

Table 2.2: Egyptian hieroglyphic numerals

1	10	100	1000	10000	100000	1000000
𓏤	𓎆	𓍢	𓆼	𓂭 𓂭	𓆐 𓆐	𓁨

The system is purely decimal (base-10) and cumulative-additive, with each sign repeated up to nine times as necessary, and ordered from highest to lowest rank. The direction in which a numeral is to be read is always the same as the direction of writing, but varies depending on the inscription in question. The set of signs in the table above are those used when the direction of writing is from left to right; when right-to-left writing is used, the signs are mirrored (i.e. 𓏤𓎆𓍢𓆼𓂭𓆐). Occasionally, when days of the month are being expressed, the signs for 1 and 10 were placed on their side: ⊂ or ═ instead of 𓎆 or 𓏤 (Gardiner 1927: 191). To aid in reading long numeral-phrases, five or more identical signs would usually be grouped in sets of three or four rather than placed on a single line. Thus, 5 is written as a row of three signs above a row of two signs, 6 as a row of three above a row of three, 7 as a row of four above a row of three, 8 as a row of four above a row of four, and 9 either as a row of five above a row of four *or* as three rows of three.[2] Given these parameters, 68257 might be expressed as shown in Figure 2.1 (reading from right to left):

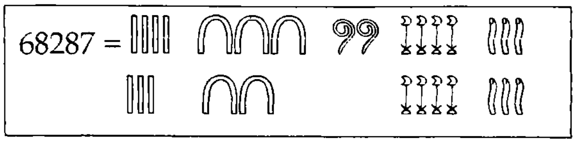


Figure 2.1: 68287 in Egyptian hieroglyphs

The sign for 1 is a simple vertical stroke. Gunn suggests that in well-executed inscriptions, the sides of the vertical bar are curved inwards slightly, thus making a biconcave bar, and postulates that it may represent "a small object of bone or wood used in some kind of tally or aid to reckoning" (Gunn 1916: 280). However, this sentiment has not been echoed by modern scholars, and I tend to think that it is simply an abstract stroke. The sign for 10 has been described as a heel bone (Kavett and Kavett 1975: 390), a tie made by bending a leaf (McLeish 1991: 42) or even anachronistically as a croquet wicket (Boyer 1959: 127). Sethe (1916: 2) correctly points out that it corresponds to the phonetic value mḏw, for a hook, handle or strap, and is thus a rebus-pictogram for the Egyptian number word for 10 (mḏw). The higher exponent signs also have specific representational qualities related to the phonetic values of the relevant Egyptian number words. The sign for 100 ( ) is probably a coiled length of rope (št), that for 1000 ( ) is certainly a lotus-plant (ḫ3), the sign for 10,000 (or) is an extended finger (ḏbʿ), and that for 100,000 ( ) is a tadpole (ḥfn). These numeral-signs, as well as the overall structure of the system, remained remarkably stable throughout its history. In some older instances in which the sign for 1000 occurs, rather than grouping the signs in clusters of three to five separated signs (as in the numeral-phrase above), multiple "lotus plants" were depicted as emerging from a single bush (e.g. 3000 =). The sign for one million (= *ḥḥ*) is only used numerically in the earliest period of the system's history. The signs for

[2] However, other groupings were sometimes used when it was more convenient for the scribe to do so.

both 100,000 and 1,000,000 originally meant "multitude", "a countless quantity" or simply "many", just as the word "myriad" can mean a group of ten thousand or more generally a large quantity (Gunn 1916). After the Early Dynastic period, the sign for 1,000,000 ceased to be used numerically, and was used only in this lexical sense. In all other respects, Predynastic numerals would have been completely intelligible to Late Period scribes.

The earliest known Egyptian hieroglyphic numerals are those from Tomb U-j at Abydos, which dates to around 3250 BC (late Naqada II or early Naqada III period), and has also produced the earliest examples of Egyptian writing (Dreyer 1998). Numeral-signs occur on a large number of drilled bone and ivory tags found in this royal tomb, which were probably once attached to containers of grave goods. Other tags have a small number of other signs that resemble later Egyptian hieroglyphs, but none contain both numerals and hieroglyphs. Some tags have 6-12 vertical or horizontal strokes, others the sign for 100, and one has both a sign for 100 and a sign for 1 (Dreyer 1998: 113-118). This system has three unusual features as compared to the mature hieroglyphic system: it uses both horizontal and vertical strokes for units, there is no attested numeral-sign for 10, and there are tags with more than nine unit-strokes. Dreyer (1998: 140) explains the first two of these irregularities simultaneously by noting that on Old Kingdom linen-lists, vertical strokes stand for 1 and horizontal strokes for 10. However, the time discrepancy of several centuries leaves me skeptical of this analogy. Furthermore, the Tomb U-j tags are very similar to others found at Naqada and Abydos that date from the Naqada III and Early Dynastic periods, which contain the sign for 10 and use vertical strokes for 1 (Dreyer 1998: 139). It is entirely possible that the writers of the Tomb U-j tags were still in the experimental stage of working with numeral-signs. The very early date of the tags makes it probable that the system was developed independently of Mesopotamian influence, although early proto-cuneiform numerals were in use around the same time.

While we have no evidence for numeral-signs higher than 100 from the Tomb U-j tags, by the Early Dynastic period the system was fully developed. The so-called Narmer mace-head found at Hierakonpolis, which may describe the unification of Upper and Lower Egypt by Narmer around 3100 BC, shows that even the very highest signs were being used at that time (Arnett 1982: 42). The mace-head indicates the booty brought back from Narmer's victorious expeditions: 400,000 bulls, 1,422,000 goats, and 120,000 prisoners (Ifrah 1985: 204). While these numbers seem grossly exaggerated, they are unambiguously numeral-signs and, in the case of the last two numbers, involve the concatenation of different exponent-signs. Another early example of hieroglyphic numerals is found on the Second Dynasty statue of Khasekhem indicating the slaughter of 47,209 of the pharaoh's enemies (Guitel 1958: 692).

In the Greco-Roman era, the hieroglyphic numerals, like the script itself, became more complex. The sign for 1,000,000 was reintroduced into the numerical sequence, though it is unclear whether its numerical meaning was truly understood. In a few inscriptions from this period, a 'ring' sign – Ω – is found in the sequence between 𓆐 and 𓁨. While Sethe (1916) believed that the ring-sign was a meaningless addition, Gunn (1916:280) protested that perhaps, in order to lengthen the series of numerals without assigning the god 𓁨 a subordinate place, Ω was assigned the value of 1,000,000 while 𓁨 either shifted upwards in value to 10 million or else retained its lexical meaning of 'an uncountable number'. Sethe's argument is strengthened by the inscription on the stele of Ptolemy Philadelphios (r. 282 - 246 BC) at Pithom, in which the sign used for 100,000 is not 𓆐 but rather 𓆐Ω, with the ring sign placed underneath the ordinary tadpole sign (Ifrah 1985: 206).

Another curious change in the Ptolemaic hieroglyphic numerals is the occasional use of cryptographic ciphered numeral-signs for many numbers, especially on the walls of the temple of Edfu (Ifrah 1998: 176-177). These signs replaced the standard cumulative sets of signs with single signs whose association with the number was homophonic, pictorial, religious, or related to the corresponding hieratic numeral-sign. They were used as early as 950 BC on a wooden votive cubit rod of Sheshonk I, but are found on no artifacts between that point and the Ptolemaic era (Priskin 2003). The most common of these signs is that for 5, a five-pointed star (𓇼), which often combines with unit-strokes in the same way that **V** = 5 in Roman numerals; thus, 9 could be written as 𓇼IIII instead of III III III (Sethe 1916: 25). However, unlike the Roman numerals and related systems, no signs were developed for 50, 500, or other half-exponents. Its origin is almost certainly pictorial, from the five points of the star. Other common signs were a human head, 𓁶, for 7, from the Egyptian understanding of the head as having seven orifices, and a scythe, 𓌳, for 9, from the resemblance between that sign and the hieratic numeral-sign for 9 (Ifrah 1998: 176-77; Sethe 1916: 25). In addition to the signs for the units 1 through 9, there were cryptographic hieroglyphs for 60, □, and 80, , both of which were derived from resemblances to hieratic numerals (Fairman 1963). These developments never resulted in the development of a fully ciphered-additive set of hieroglyphic numeral-signs, since these signs were often included in otherwise perfectly normal cumulative numeral-phrases.

Hieroglyphic inscriptions are largely monumental and are written on stone. Texts including hieroglyphic numerals include seals, funerary stelae and tomb-inscriptions, annals, lists relating to conquest and plundered goods, and certain administrative texts. An often-overlooked source of hieroglyphic numerals is the wide variety of stone balance-weights bearing inscriptions indicative of their weight (Petrie 1926; Petruso 1981). Numerals were used to indicate dates, weights and measures and, of course, a wide variety of cardinal quantities of goods, animals, and people. In all of these texts, the numerals are usually formed in the ordinary fashion described above.

An exception to the cumulative-additive use of hieroglyphic numerals is found on the Palermo Stone (Vth Dynasty, ca. 2400 BC), perhaps the most famous of the early Egyptian pharaonic annals. It is an annal of the activities and events concerning the pharaohs, but also contains data such as the height of the Nile at full flood and surveys of areas of land. The Palermo Stone annalist uses an unusual means of expressing quantities of the *aroura* measure of land (Clagett 1989: 56-57). Normally, one would place a numeral-phrase after the metrological sign for 1, 10, or 100 arouras to indicate the total quantity. However, in the modified aroura-system, the metrological sign for the highest aroura-value was omitted, and only the units for that exponent were marked. Thus, instead of writing [10-aroura glyph] || [1-aroura glyph] ||| for 23 arouras, the scribe could omit the 10-aroura sign. No information would be lost; because the numerals must always follow the metrological sign, that || means 2 x 10 arouras is the only possible interpretation. This sort of usage resembles a cumulative-positional numerical notation system, in that, instead of writing the highest metrological sign, it is omitted and its value is to be understood by its position. Nevertheless, Clagett admits that this system is not used regularly throughout the Palermo Stone and is not found in any other inscriptions; hence its value for understanding the hieroglyphic numerals is somewhat limited (Clagett 1989: 57).

One purpose for which the hieroglyphic script was definitely not used was for mathematics and calculation. The vast majority of Egyptian literature, and all of the mathematical texts of Egypt, are written in the hieratic or later demotic scripts (cf. Gillings 1978: 704-5). Nor do we find hieroglyphic numerals marked on potsherds, tallies, or other such media that would suggest their use as an intermediate step in performing calculations. We do have some hieroglyphic evidence indicating the calculation of the area of a rectangle from the inscription from the tomb of Methen (IVth Dynasty, 26th century BC), but this inscription simply indicates that the calculation *was done*; the numerals were not actually used in the calculation process (Peet 1923: 9).

A significant problem in the history of Egyptian numeration has arisen because Egyptologists regularly transliterate documents in the hieratic script (which varies tremendously depending on the time period of the text and the idiosyncrasies of the scribe's handwriting) into regularized hieroglyphs. For most paleographic matters, this convention does not present a great problem, but the hieroglyphic and hieratic numerals (which will be discussed later in this chapter) employ very different principles: the number 999 requires 27 separate hieroglyphic signs but only 3 hieratic signs, as shown in Figure 2.2.

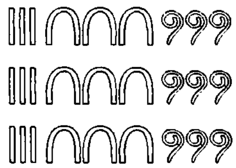
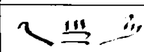
	Hieroglyphic	**Hieratic**
999		

Figure 2.2: 999 in hieroglyphic and hieratic

Egyptologists have long been aware of this contrast in structure, but have not always expressed this awareness in their writings. The Egyptologist T. Eric Peet, one of the earliest experts on Egyptian mathematics, neglected to mention the hieratic numerals in his summary of Egyptian mathematical capabilities (Peet 1931: 411), although he made it clear in his classic transcription and translation of the famed Rhind Mathematical Papyrus that hieratic numeral-phrases are often far shorter than their hieroglyphic counterparts (1923: 11). In the same way, historians of mathematics regularly fail to consider the different principles by which the two systems are structured, and use this 'fact' to criticize Egyptian mathematics as cumbersome and clumsy. Over forty years ago, Carl Boyer, the pre-eminent historian of mathematics at the time, decried the failure of historians of mathematics to recognize that while hieroglyphic numerals are not very concise and would be difficult to use for mathematics, the hieratic numerals do not share this deficiency and are in fact quite easy to work with (Boyer 1959). In fact, many historians of mathematics do not mention the hieratic (or the later demotic) numerals at all. Unfortunately, the great number of recent scholars who continue to write as if the hieroglyphic numerals were the only ones available to Egyptian scribes suggests that this point needs to be re-emphasized (cf. McLeish 1991: 42; Guedj 1996: 34-5; Palter 1996: 228-229; Dehaene 1997: 97). It is necessary to treat the hieroglyphic and hieratic systems separately, not despite their very strong historical connection, but *because* of that connection, inasmuch as the two systems were different in structure and used in entirely different functional contexts. It is entirely inappropriate to discuss the mathematical efficiency of hieroglyphic numerals when they were never used for mathematics.

On a very limited number of hieroglyphic documents – only two that we know of - large numbers (particularly multiples of 100,000) were expressed through multiplicative formations instead of purely additive ones. For instance, in one text dating to the Ptolemaic era, the number 27,000,000 is expressed as shown in Figure 2.3, using a single sign for 100,000 underneath which the ordinary additive hieroglyphic phrase for 270 was written (Sethe 1916: 9),

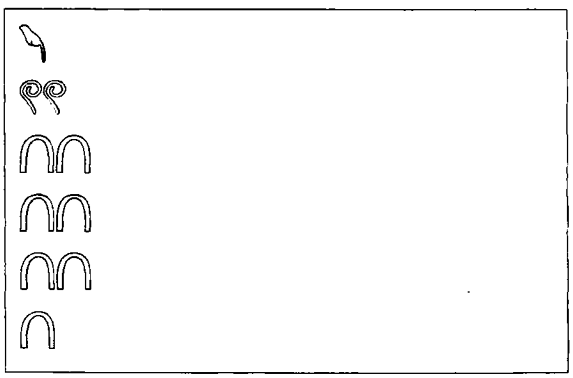


Figure 2.3: Multiplicative phrase for 27,000,000

Because the sign for 1,000,000 was only used in early periods of Egyptian history, the only other way to express the number 27,000,000 would have been to use 270 signs for 100,000. There is thus a clear economy in using the multiplicative principle instead of pure cumulative-additive structuring. In a second instance (from the time of Amenhotep III, around 1400 BC), curiously, 100,000 is expressed multiplicatively using the tadpole-sign  placed above a vertical stroke – thus, 100,000 x 1 = 100,000 (Sethe 1916: 9; Loprieno, personal communication). This example is particularly interesting in that, while it uses multiplication, it does not contribute to the economy of signs used, but rather increases the number of signs needed from one (the tadpole-sign alone) to two (tadpole + vertical stroke). In these two numeral-phrases, the hieroglyphic system is thus not purely cumulative-additive but is a hybrid that is multiplicative-additive above 100,000.

Despite these two examples, I do not find the evidence for widespread hieroglyphic multiplicative notation to be persuasive. The ciphered-additive hieratic numerals use multiplicative forms far more frequently than do the hieroglyphs. Sethe (1916), whose study of the topic is unfortunately the most recent available, lists only two hieroglyphic examples (the ones mentioned above) as opposed to more than ten hieratic examples, with others mentioned by other authors (cf. Möller 1927: 59). Unfortunately, it is the hieratic examples that have been mentioned in Egyptian grammars such as Gardiner (1927: 191), particularly those numerals found on the important Harris and Kahun papyri. Coupled with the unfortunate tendency to transliterate hieratic numerals as hieroglyphic numerals, this creates a serious problem. Many historians of mathematics (and some Egyptologists) have concluded that multiplicative expressions are common in the hieroglyphic numerals, when in fact almost all such expressions come from hieratic texts.

The question of the origin of multiplicative structuring in the hieroglyphic numerals has not been satisfactorily answered. Gardiner's (1927: 191) suggestion that the adoption of multiplicative hieroglyphic forms was a consequence of the loss of the sign for 1,000,000 is on track but is too simplistic. I think it probable that the Egyptian scribes, having developed the hieratic script and numerals as a cursive shorthand based on, but quite different from, the hieroglyphs, recognized the benefit of increased conciseness in the specific case of multiplication by 100,000 and borrowed it for the hieroglyphs as well. Hieratic multiplicative forms were being used already in the Middle Kingdom (the Harris and Kahun papyri are probably both from the Twelfth Dynasty), whereas the only two known examples of hieroglyphic multiplication are from the New Kingdom and the Ptolemaic era. All of the examples of hieratic numerals greater than 100,000 discussed by

Möller (1927) are expressed in this multiplicative fashion. We thus find in the Egyptian hieroglyphic multiplicative numerals a case where an ancestral system (hieroglyphic) borrows a structural feature from its descendant (hieratic). Because hieroglyphic numerals are only used for monumental purposes, numbers higher than 100,000 would have been expressed only infrequently, as opposed to practices associated with the hieratic script, which was used for most administrative and mathematical purposes. It is entirely possible, given the small number of hieroglyphic inscriptions using this structure, that it was not part of the standard scribal education but rather an exceptional response to the occasional requirement for expressing high numbers in hieroglyphic numerals. The question of when this borrowing took place remains open; the first hieratic documents to use this structure date to the Middle Kingdom (2040 to 1652 BC), while the first hieroglyphic example (mentioned above) dates to about 1400 BC.

Guitel regarded this development in the structure of the hieroglyphic numerals as being particularly important because a small number of Aztec numeral-phrases (see ch. 9) show a similar development (Guitel 1958: 692-695, 1975: 70-73). The two systems are similar in structure, the main difference between the two being that the Egyptian hieroglyphs have a base of 10 where the Aztec numerals have a base of 20. The independent development of this feature in these two cases is notable. However, I cannot agree with Guitel's argument that these multiplicative formations represent a step (or even an abortive step) towards a fully positional notation (Guitel 1975: 44). Rather, it represents an alternative means of increasing the conciseness of some (but not all) numeral-phrases and extending a system's capacity to write numbers while retaining its basic structure.

In addition to the system for representing whole numbers, the Egyptian hieroglyphic script possessed two distinct systems for representing fractional values. Both of these systems could express only *unit-fractions* – those in which the numerator is 1. The first such system, the standard system for expressing fractional quantities, simply required the scribe to place the "mouth"-sign ( ), which also meant "part", above any hieroglyphic numeral-phrase to indicate the corresponding unit fraction (Loprieno 1986: 1307). Alternately, if the mouth-sign was too small to place over the entire phrase, it was simply placed over the signs for the highest exponent of the denominator, with the other signs placed after it. The numbers 1/12 and 1/246 are depicted in Figure 2.4.

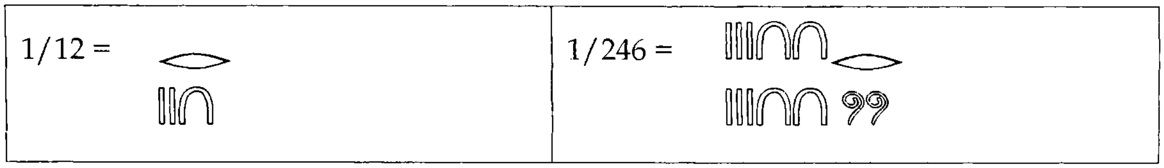


Figure 2.4: Hieroglyphic fractions for 1/12 and 1/246

This system also used three special symbols for some of the most commonly used fractions: 1/2= , 2/3 = , and 3/4= (Sethe 1916: Table II). The last two of these are not unit-fractions, and are the only exceptions to the general rule that all fractions must be written as unit fractions. This system was not used in the Predynastic era, but is certainly found in abundance during the Old Kingdom and thereafter. While, in theory, any fraction could be expressed with this system, the majority of Egyptian hieroglyphic fractions are larger than 1/20.

The second system was used only for measurements of volume of grain, fruit, and liquids by indicating fractions of the *hekat* (ḥḳ3t), a measure probably equal to 4.8 litres (Ifrah 1985: 208). This notation is known as "Horus-eye fractions" because the various symbols for fractional values can be combined to form the glyph of the Wadjat (wḏ3t) or eye of Horus ( ), a symbol of health, fertility, and abundance. These signs are shown in Table 2.3, presuming a left-right direction of writing (Sethe 1916: Table II).

Table 2.3: 'Horus-eye' fractions

1/2	
1/4	
1/8	

1/16	
1/32	
1/64	

These signs can be put together to form the Horus-eye; however, the sum of these signs is only 63/64. Since the value of each successive sign is exactly one-half of the preceding one, the totality - one - cannot ever be reached. This system for expressing fractions is essentially binary and bears no relation to either the standard hieroglyphic numerals or the ordinary unit-fraction system. It is probable that its binary structure is relevant insofar as dividing and multiplying by two is a standard operation needed when manipulating volumes of goods. This system may have originated in an earlier *hieratic* series of fractional signs, of which the earliest example is from the Abusir Papyri of the Fifth Dynasty (Reineke 1992: 204). It appears that only later did the signs become assimilated to the parts of the Horus-eye symbol, as the first hieroglyphic Horus-eye fractions are from the New Kingdom (Priskin 2002: 76).

The Egyptian hieroglyphic numerical notation system has a number of direct descendants. As mentioned earlier, its most direct descendant is the Egyptian hieratic system, which developed as early as the First Dynasty as a scribal shorthand for the hieroglyphs (Peet 1923: 11). Egyptian scribes certainly would have learned both the hieroglyphic and hieratic numerals during their education, and used both systems in the appropriate contexts - the hieroglyphs engraved on stone monuments, and the hieratic numerals written in ink on papyrus and ostraca. However, the hieratic numerals are ciphered-additive rather than cumulative-additive, and thus embody a significant structural change.

It is also very likely that the civilizations of the Aegean used the Egyptian hieroglyphic numerals as the model for their own indigenous numerals - the Cretan hieroglyphic system and the Linear A and B numerals. There was considerable commercial and political interaction between Egypt and the Aegean in the early 2nd millennium BC, just at the time when the Aegean numerical notation systems began to emerge. Despite the dissimilarity in the forms for the numeral-signs of the two systems, they are identically structured, and thus it can be safely asserted that the Minoans did not develop their numerals entirely independently, but instead borrowed at least the idea of the numerals from the Egyptians. The exact kinship of the Hittite hieroglyphic numerals system to the Egyptian hieroglyphs and the systems of the Aegean is unclear, but it is possible that it was developed directly on the Egyptian model around 1400 BC.

Another system that is more or less directly descended from the Egyptian hieroglyphs is the early Phoenician-Aramaic one, which began to be used around 750 BC in the Levant. The situation is made more complex, however, since the Phoenician-Aramaic system blends the form of the numeral-signs and the direction of writing of the Egyptian hieroglyphs with the structure of the Assyro-Babylonian common (decimal) numerals. Alternately, it is also possible that the Phoenician-Aramaic system was based on the Hittite hieroglyphs rather than the Egyptian numerals, in which case its debt to the latter was indirect. This development marks the formation of the Levantine family of numerical notation systems (ch. 3), and reflects the intermediary position of the Levantine civilizations between two larger polities of the 8th century BC – the Egyptian and Assyrian states.

By the Greco-Roman period, the use of the hieroglyphic script and numerals declined greatly, and both script and numeration increased in the number of signs used and the complexity thereof, to the point where it was considered to be a purely symbolic or cryptographic script by outsiders (Ritner 1996: 81). By the 3rd century AD, Egypt was becoming increasingly Christian in its religion, and its language was being written in the Greek and Coptic scripts. The latest dated hieroglyphic inscription dates from 394 AD, and by the fifth century, knowledge of how to read and write hieroglyphs had disappeared. The hieroglyphic numerals were the last of the Egyptian systems to go extinct, and were replaced by the Coptic alphabetic numerals (ch. 5).

Egyptian Hieratic

The hieratic script was developed around 2600 BC by Egyptian scribes as a sort of cursive shorthand for the earlier hieroglyphic script, and continued to be used in some parts of Egypt as late as 200 AD (Loprieno 1995). Like its forerunner, hieratic was a mixture of logographic and phonographic components. However, unlike the hieroglyphs, which were usually written on stone, hieratic was designed for cursive writing on papyrus and on ostraca, making it suitable for administrative and literary purposes. Furthermore, while the hieroglyphs could be written in a variety of directions, hieratic texts are always linear and written from right to left. While the form of the hieroglyphs was very regular and formalized, hieratic writing varied greatly by period, location, and the idiosyncrasies of the scribe's handwriting. The Old Kingdom divergence of Egyptian scripts into monumental (hieroglyphic) and cursive (hieratic) variants produced a dualism that continued throughout the remainder of ancient Egyptian history.

A base-10 ciphered-additive numerical notation system accompanied the hieratic script. The hieratic numeral-signs, like the script itself, changed considerably over the system's extensive history. The paleographic development of hieratic numerals is traced in the charts provided by Möller (1936). In Tables 2.10, 2.11, and 2.12, I present three distinct sets of numerals, the first and earliest from the Kahun papyrus, from the Twelfth Dynasty (20th and 19th centuries BC), the second from Pap. Louvre 3226 (15th century BC), and the third from the Harris papyri, (12th century BC) (Möller 1936, vol 1:59-63, vol. 2: 55-59). I use these three because they contain mostly complete sets of numeral-signs at least as high as 1000, and are thus very useful for comparative purposes, even though they reflect only a limited portion of the system's history. The Harris papyri numerals, from the second table below, include all of the cardinal numbers up to 100,000; this is the only text to do so.

Table 2.10: Hieratic numerals (Kahun papyrus, Twelfth Dynasty)

	1	2	3	4	5	6	7	8	9
1s									
10s									
100s									
1000s									
10,000s									
100,000s									

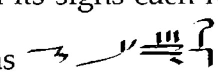
Table 2.11: Hieratic numerals (Pap. Louvre 3226, Eighteenth Dynasty)

	1	2	3	4	5	6	7	8	9
1s									
10s									
100s									
1000s									

Table 2.12: Hieratic Numerals (Pap. Harris, Twentieth Dynasty)

	1	2	3	4	5	6	7	8	9
1s									
10s									
100s									
1000s									
10,000s									
100,000s									

We can see considerable change in the numeral-signs by comparing these sets of numerals. Looking only at the signs for 5, 6, 7, and 9, the three series appear remarkably distinct. At the same time, however, a large majority of the hieratic numeral-signs show remarkable continuity. Many of the hieratic signs used in the Old Kingdom would have been perfectly comprehensible to a scribe in the Late Period or even the Ptolemaic era. Note, however, that many of the numeral-signs are very similar to others from the same period; for instance, it is very difficult to distinguish 400 from 600 or 3000 from 5000 in Table 2.12. When used to express days of the month, hieratic numerals, like hieroglyphic numerals, were often rotated 90 degrees counter-clockwise to reflect this separate function. Given the nature of the Egyptian calendar, these forms only exist for numerals less than 30. To write fractional values, a small dot was placed above the numeral-phrase for an integer to indicate the appropriate unit fraction (1/x).

The hieratic system is primarily ciphered-additive and its signs each represent a multiple of an exponent of 10. Thus, 56207 could be written as . Many of the hieratic numeral-signs bear a clear relationship to their cumulative-additive hieroglyphic forerunners, which is particularly evident in the signs for 1 through 4, 10, 10,000 through 40,000, and 100,000 (cf. hieroglyphic 𓏺, 𓏻, 𓏼, 𓏽, 𓎆, 𓂭, 𓂮, 𓂯, 𓂰, and 𓆐). Other hieratic numerals show no clear correspondence with their hieroglyphic ancestors except in very early periods. The hieratic numerical notation system is interesting because it is primarily a ciphered-additive system, but clearly shows its cumulative-additive ancestry. For this reason, I include the hieratic system and its immediate descendants in this family even though it is very different from its hieroglyphic ancestor.

For writing many of the values above 10,000, the Egyptians used multiplicative notation; for instance, the sign for 60,000 is written by placing the sign for 6 below the sign for 10,000. This principle is not used for 10,000-30,000, but was used occasionally for 40,000 (as can be seen in the Kahun numeral-sign), and was regularly employed for

50,000-90,000 and for values above 100,000. While it may appear that the multiplicative principle is used for certain values of the hundreds and the thousands, this is not the case. A careful paleographic analysis of the development of the numeral-signs shows that 300 ( ) developed out of an earlier form, 𓍢𓍢𓍢, and thus represents the slow abbreviation of the first two of three cumulative 100-signs and the extension of the third rather than the juxtaposition of 3 and 100. Möller lists only one occurrence of a hieratic sign for 1,000,000 to correspond with the hieroglyphic 𓁨, but as with its hieroglyphic counterpart, this hieratic numeral-sign appears to have been abandoned after the Old Kingdom. Later, the regular use of multiplicative-additive structuring allowed numbers above 100,000 to be expressed easily in hieratic by placing the appropriate multiplier below the 'tadpole' sign. The first such instances listed by Möller are in the Kahun papyrus, where 200,000, 500,000 and 700,000 are expressed by the "tadpole" sign for 100,000 placed above the appropriate unit-signs, so I tentatively date the origin of this principle to the 20th century BC. While, as discussed above, there are only two known examples of hieroglyphic multiplicative numerals, many texts include hieratic multiplicative ones. This suggests that the development of hieratic numerals was not simply as a scribal shorthand for hieroglyphs, but a highly creative process involving not only the shift from cumulative-additive to ciphered-additive notation, but also the use of multiplication where it was deemed useful for abbreviatory purposes.

The strong similarities between the hieratic numerals and the earlier hieroglyphic numerals, coupled with the indisputable historical connections between the two scripts, indicate the historical indebtedness of the hieratic to the hieroglyphic numerals. Whereas the hieroglyphic numerals are found in Predynastic inscriptions, hieratic numerals first appear in the First Dynasty (Peet 1923: 11). Their use became widespread from the Old Kingdom onwards, with the two systems (hieroglyphic and hieratic) being used for parallel purposes. There can have been no influence on the hieratic system from systems other than the hieroglyphic, because the only other numerical notation system in use at this period was the Sumerian cuneiform system, which is entirely unlike it.

The development of hieratic thus represents the first step towards ciphered notation in the history of numerical notation, and an important step away from the use of one-to-one correspondence between signs and their signifiers. However, this invention did not take place at a single point in time. The Old Kingdom hieratic numerals were little more than cumulative-additive cursive forms of the appropriate hieroglyphic numerals. Over time, the numerals became increasingly removed from their hieroglyphic ancestors as multiple strokes were condensed into single strokes, probably for greater ease of writing. Table 2.13 compares the way in which the numbers 5 through 9 and 300 were written in Old Kingdom hieratic to the numeral-signs from the three sets of numerals presented above:

Table 2.13: Evolution of cursive from linear Egyptian numerals

	Hieroglyphic	Old Kingdom	Kahun papyrus (Dyn. 12)	P. Louvre 3226 (Dyn.18)	P. Harris (Dyn. 20)
5					
6					
7					
8					
9					
300					

While ciphered signs were the ordinary ones, the system's evolutionary origins were not completely forgotten; the cumulative numeral-signs were occasionally employed even into the New Kingdom. It is probable that no single individual invented ciphered notation in Egypt, but rather that its development was a process of abbreviating and combining cumulative signs by scribes over many centuries, until, by the Late period, very few hieratic signs bore any resemblance to their hieroglyphic counterparts. It is even possible that the scribes making these changes were not really aware of the importance of the new structural principle they were using. Hence the origin of ciphered notation may, in some sense, have been accidental.

An even more remarkable development in some hieratic documents from the Ptolemaic era is that there is a reversion in the numeral-signs away from the ciphered signs used in older hieratic texts as well as from the demotic numerals more common at that time, and a returning to the common use of the cumulative principle. For instance, Möller lists several texts (Leinwand, P. Bremner, Isis-N., Leiden J. 32, and P. Rhind[3]) in which hieratic units were expressed with repeated vertical strokes (𓏺), tens with horseshoe-shaped curves (𓎆), and hundreds with coils (𓍢), in an exact imitation of the hieroglyphic numeral-phrases of the same value (Möller 1936: vol. III, 59-60). While some of these documents retained the ciphered signs for some values, there is an obvious trend over time towards the use of cumulative numeral-signs in hieratic documents. It is possible the scribes in these cases had forgotten the ciphered signs; however, the existence of demotic ciphered-additive numerals at that period makes this possibility unlikely. Rather, the reversion to cumulative-additive numerals in hieratic texts was probably a deliberate archaism, resulting from the desire to emulate hieroglyphs more exactly. To my knowledge, no Egyptologists have studied this topic or remarked on this change in structural principle, in part, no doubt, because the difference in principle

[3] P. Rhind does not refer here to the famous Rhind Mathematical Papyrus, but to a different text dating to 9 BC and having nothing to do with mathematics.

between the hieroglyphic and hieratic numerals remains poorly recognized among both Egyptologists and historians of mathematics (cf. Boyer 1959).

Egyptian scribes would have learned both hieroglyphic and hieratic writing and numerals during their education, and used whichever was appropriate according to the context - hieroglyphs on stone monuments, hieratic on papyrus and potsherds. Accordingly, while the functions of the hieratic numerals are quite distinct from those of the hieroglyphic numerals, the users of the two systems would have been the same individuals. For the hieratic numerals, two functions stand out above all others: administration and mathematics.

Hieratic numerals were used with overwhelming frequency for administrative purposes throughout the history of ancient Egypt. They are found on a variety of papyrus documents and ostraca throughout their history. Almost all extant Egyptian legal, commercial, educational, and literary texts from 2600 to 600 BC are written in hieratic, and numerals abound on such documents. While hieroglyphic numeral-phrases were very lengthy, requiring an enormous number of symbols to express many small values, hieratic numerals were highly concise, enabling their use in accounting, commerce, and law, as well as expressing dates and cardinal quantities. Because they would have been learned and used by only a small and well-educated segment of the populace (i.e. the scribes), their main disadvantage - the large number of signs one needed to learn in order to use the system - would not have been a serious problem. Learning the numerals would have been an early part of scribal education.

However, most academic discussion of the hieratic numerals over the past century has focused not on their administrative functions, but rather on a limited, if very interesting, set of texts dealing with mathematics. The hieratic numerals were the first ones to be used in Egypt for arithmetic and mathematics in the late Middle Kingdom and the early Second Intermediate Period (XIIth and XIIIth Dynasties). The Reisner, Berlin, Kahun, and Moscow mathematical papyri are among these early texts, all dating from the

19th century BC. Later, around 1650 BC, during the period of Hyksos domination, the so-called Egyptian Mathematical Leather Roll and the famed Rhind Mathematical Papyrus were written using hieratic numerals, though the latter may be a copy of an earlier document. All these remarkable texts show that it is entirely possible to perform arithmetic and simple mathematics with hieratic numerals. However, other than from the brief period of Egyptian history from roughly 1900 to 1650 BC, we have no evidence that the hieratic numerals were used for purposes which mathematicians today would consider part of their science. A full discussion of the mathematics of ancient Egypt is well beyond the scope of this work, but many good discussions are available for the interested reader (Peet 1923; Neugebauer 1957; van der Waerden 1963; Gillings 1972, 1978).

It is probably fair to say that the mathematics of the Egyptians could not have been done with the hieroglyphic numerals alone, because the older, cumulative-additive system is not very concise. On the other hand, the hieratic numerals, being essentially ciphered-additive, were no different in structure from their eventual descendant, the Greek alphabetic numerals, for which we have abundant evidence of their mathematical use. However, to focus too closely on the strictly mathematical functions of the hieratic numerals would be a grave error. Regardless of their interest from a mathematical perspective, the six documents mentioned above are but a minuscule fraction of the total number of hieratic texts containing numerals. From a paleographical or historical perspective, the non-mathematical hieratic texts give us a much better idea of their function and structure than do the mathematical ones. We will probably learn much more about the arithmetical efficiency of hieratic numerals from their use in bookkeeping and administration than from mathematical texts. Unfortunately, because Egyptologists have paid little attention to the hieratic numerals over the past half-century, and because historians of mathematics have focused their attention quite naturally on the mathematical texts alone, the body of research needed to fully understand their use does not exist at present.

Tracing the diffusion of the hieratic numerals is quite difficult. As the system was primarily used for administrative purposes, it spread wherever Egyptian domination extended, for instance, into Canaan in the nineteenth and twentieth Dynasties (Millard 1995: 189-90). The early Israelites used a variety of hieratic numerals (described below) starting in the 10th century BC. In addition to their geographical diffusion throughout the Egyptian world, the hieratic numerals gave rise to two distinct descendant systems. First, the demotic numerical notation developed out of the hieratic starting in the 8th century BC, and eventually came to replace its ancestor. In addition, the Meroitic cursive script, found on ostraca in the Sudan starting in the 3rd century BC, contains numeral-signs to which Griffith (1916: 23) assigns ancestry from the hieratic numerals, although, as I will show below, this simple derivation is not without problems. While the hieratic numerals have relatively few direct descendants, through its demotic descendant, they are ancestral to a great number of systems.

In the Twenty-sixth Dynasty (664 to 525 BC), the demotic script and numerals, which had only begun to diverge from hieratic a century or so earlier, were accorded royal preference for most purposes. After that point, demotic began to replace hieratic for more and more functions throughout Egypt. By the early Christian era, when hieratic was encountered by the Greeks, it was only used in religious texts – by which means it got its name, *hieratikos* "sacred". It is ironic that the name that we now give to this script and numerical notation system is taken from a purpose for which it was rarely used throughout over two millennia of its history. By around 200 AD, even these religious functions had ceased.

Hebrew Hieratic

Around the 10th century BC, the Egyptian scribal tradition, including the use of the hieratic script and numerals, came to be used by the ancient Hebrews, around which time the Israelites incorporated a great deal of Egyptian learning into their own thought. Prior to this point, there is no evidence that the Israelites used any numerical notation whatsoever, although it is likely that many would have become familiar with Egyptian notations while in that land.

Our best evidence for the use of hieratic numerals among Hebrew scribes comes from a large ostracon found at Tell el-Qudeirat (Kadesh-barnea) in 1979, upon which is found a very complete series of numerals; only the signs for the units and 60 were missing, blurred or unreadable (Cohen 1981: 105-107). These are shown in Table 2.14.

Table 2.14: Hebrew hieratic numerals

	1	2	3	4	5	6	7	8	9
10s									
100s									
1000s									

The Kadesh-barnea ostracon dates to the 10th century BC, and was probably a scribal exercise in writing numerals and measures. We can be certain of its attribution to Hebrew-speaking peoples because of the presence of Hebrew words indicating units of measurement, so it cannot possibly have been written by a foreign (i.e. Egyptian) scribe. Nevertheless, the numeral-signs are paleographically very similar and structurally identical to the late hieratic ones, so it is obvious that these Hebrew numerals were directly borrowed under conditions of political domination by and cultural contact with Egypt.

Ostraca from a later date, such as the famed Samaria ones from the first half of the 8th century BC, help fill in the units and confirm the use of hieratic numerals as a regular practice in Israel. These ostraca, found at Arad and Lakish, record weights and measures, and probably served an administrative function. While Gandz (1933: 61) argued that the numerals from the Samaria ostraca had an Aramaic origin, an examination of the numeral-signs in question indicates clearly that the numerals are hieratic, as shown in Table 2.15 (cf. Lemaire 1977: 281).

Table 2.15: Hebrew hieratic, Egyptian hieratic and Aramaic numerals

	1	2	3	4	5	10	20
Samaria ostraca	I	II	III	IIII	⌉	⋏	⋠
Late Hieratic	I	II	III	IIII	⌉	Λ	Λ̊
Aramaic	I	II	III	IIII	III II	¬	=¬

It is difficult to know what to make of the claim (now largely abandoned) that the Hebrews used a cumulative-additive linear numerical notation system unrelated to the various Egyptian numerals. Allrik, using discrepancies between the biblical lists of Israelites in the books of Nehemiah and Ezra, concludes that many of these discrepancies can be explained by arithmetical errors made by the books' individual compilers, and that these errors indicated the use of a base-10 cumulative-additive system with a special sign for 5 (Allrik 1954). Similarly, Yadin (1961) insists from metrological evidence that the numbers inscribed on Hebrew shekel-weights could not be hieratic, but by virtue of their relative weights must represent an entirely independent cumulative-additive numerical notation system, having a decimal base but with special signs for 4, 5 and 8, as indicated in Table 2.16.

Table 2.16: Shekel-weight numerals

1	2	3	4	5	8	10
I	II	III	⌉	Λ	T	—

Yadin argues that because many four-shekel weights are marked ⌉, that sign must mean 4, and similarly that the "T" sign found on eight-shekel weights must mean 8

(1961: 21). This evidence is indeed compelling. However, Aharoni (1966) has provided a solution that reconciles the similarity of the numeral-signs to hieratic numerals with the metrological evidence. Aharoni points out that the Egyptian *deben* weight is equal to eight shekels, and that one *deben* is equal to ten *qedet*. It is thus entirely conceivable that four-shekel weights would be marked (meaning 5!) to indicate 5 *qedet* (equal to four shekels), and similarly, 10, 20, and 30 would be marked on weights measuring 8, 16, and 24 shekels, respectively (Aharoni 1966: 18). Kaufman (1967) argues that Israelite merchants and scribes would have read the hieratic numerals at face value (5, 10, 20, 30) rather than as multiples of shekels (4, 8, 16, 24). Thus, it is clear that these weights were inscribed and read with easily recognizable variants of the late hieratic numerals, further confirming the close relationship between Israel and Egypt in both numeration and metrology.

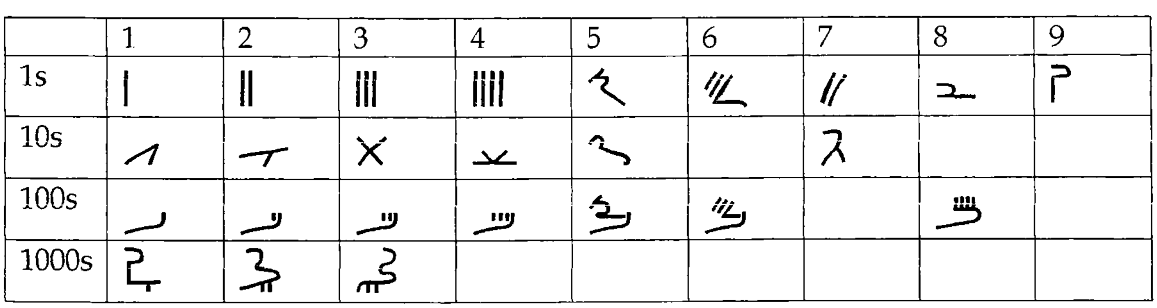
The Kadesh-barnea and Samaria ostraca, as well as the shekel-weights, are of a relatively early date in the history of Hebrew writing. After this point, we have little evidence for the use of any numeral system, hieratic or otherwise, until around 450 BC, at which time the Aramaic numerals (ch. 3) were occasionally used among Hebrew speakers, especially in Egypt. The Hebrew variant of the hieratic numerals is not directly related to the Levantine systems (ch. 3), which are cumulative-additive and have other unusual structural features. Thus, the systems used by Aramaeans, Phoenicians, and other neighbours of the Israelites were entirely different from it. There is no evidence of a distinctly Hebrew numerical notation system until about 125 BC, when the use of the familiar alphabetic numerals (ch. 5) began.

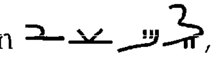
Meroitic

The kingdom of Meroë, which flourished from roughly 300 BC to 350 AD, made use of two distinct scripts. The first, Meroitic hieroglyphs, were based on Egyptian hieroglyphs and were used on some stone monuments. No numerical signs are present in the small corpus of Meroitic hieroglyphic inscriptions, although the surviving evidence is insufficient to assert that such signs were never used. The other script, the Meroitic cursive, was written from right to left on ostraca as well as on stone. The cursive script did make use of written numerals, although, as we shall see, their interpretation is still incomplete. Almost all of our information on the Meroitic numerals rests on the work of F. Ll. Griffith, the original decipherer of the Meroitic scripts. Unfortunately, because the Meroitic language has no known relatives, we are largely unable to read Meroitic inscriptions, even though the values for the signs of the cursive script are more or less fully deciphered.

Griffith (1916: 22) offered the interpretation of the Meroitic numeral-signs shown in Table 2.17, which he considered tentative, but which has not been contested since that time. However, on structural and paleographic grounds, the values for the units, 10, and all of the hundreds are unquestionable, and the remainder of the numeral-signs are fairly certain.

Table 2.17: Meroitic numerals



This system is ciphered-additive and decimal, and written from right to left (like the Meroitic cursive script). Thus, 2348 would be written , and, in fact, this number appears on the stela of Akinidad (Griffith 1916: 22). As with the hieratic numerals, there is some evidence for the use of cumulative notation in the signs for the units and low hundreds, and possibly 1000-3000 as well. There is only one case (again, from the Akinidad stela) where a number greater than 10,000 is expressed; interestingly, where hieratic uses a single sign for 10,000 ( ), Meroitic appears to use a multiplicative formation (10 x 1000). However, this evidence is far too limited to conclude that the Meroites regularly used multiplicative-additive structuring to express higher exponents.

Further evidence for the relative independence of Meroitic from the Egyptian numerical notation systems is the interesting fractional system used on many ostraca (Griffith 1916, 1925). As discussed earlier, hieratic and demotic use ciphered fractions whose signs are either independent of the relevant integer or are made by placing a dot above the integer. However, Meroitic fractional numerals were written with cumulative notation using dots ( ) to indicate tenths of the unit, and a dot in a semicircle ( ) for one-twentieth of a unit. This fractional system was clearly used in conjunction with the system for integers, rather than parallel to it; thus, 19.3 is indicated on one ostracon by the numeral-phrase (Griffith 1925: Plate XXVI). Given the relatively undeciphered nature of the Meroitic scripts, it is impossible at present to know whether these tenths could be used as abstract numbers or whether, as Griffith believed, they only represented a metrological value of one-tenth of some larger unit (1916: 22-23). Regardless of their function, this system is unlike any of the standard Egyptian systems for expressing fractional values.

By the time of the development of the Meroitic scripts, the hieratic script and numerical notation system had largely been replaced by the demotic throughout Egypt. Nevertheless, on paleographic grounds (especially the signs for 6, 10 and 20, but also the cumulative unit-signs), Griffith argued that the Meroitic numerals were more similar to the late hieratic numerals (8th to 3rd centuries BC) than they were to any of the demotic forms (1916: 23). If correct, this insight is particularly interesting, inasmuch as the characters of the Meroitic cursive script are almost certainly derived from a demotic rather than a hieratic prototype (Millet 1996: 85). The paleographic evidence is not firm enough to decide which system was the immediate ancestor of the Meroitic numerals, but that they have an Egyptian cursive origin can hardly be questioned.

Due to our inability to interpret the inscriptions, the functions of the Meroitic numerals remain obscure. From their use on both stelae and ostraca, it is likely that they were used for administrative purposes such as taxation and mensuration, as well as in funerary and monumental contexts indicating year-dates and quantities of individuals. Griffith suggests that something akin to the Egyptian *heqt* or *artaba* measures, used to indicate volumes of produce such as corn or dates, was probably indicated on some ostraca (1916: 23). There is limited evidence that the Meroitic numerals were ever used for arithmetic and none for their mathematical use. Even on ostraca upon which multiple numerals have been written, Griffith was unable, except in one instance, to establish any arithmetical correspondence between the numerals that would indicate a tally or sum had been taken (1916: 24).

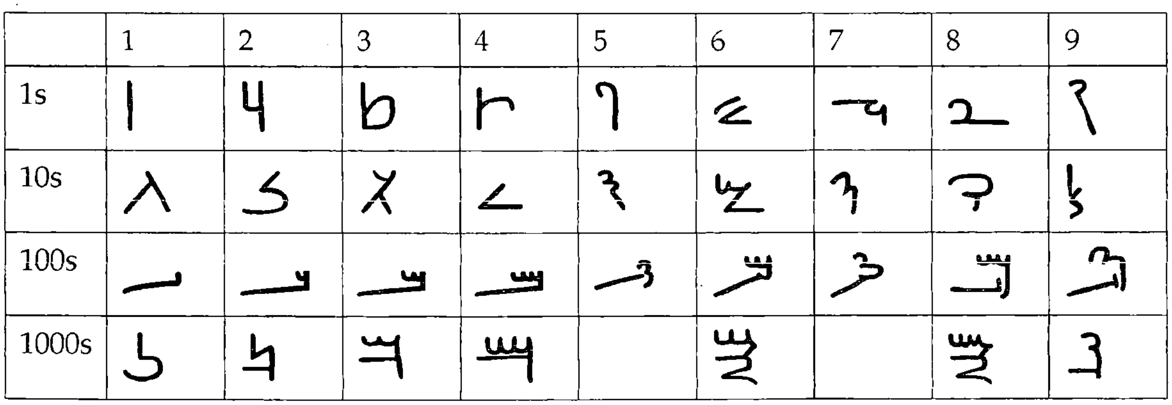
The Meroitic numerals were used until the 4th century AD, but did not outlast the kingdom of Meroë. Millet (1996: 84) suggests that the script may have continued in use until the introduction of Coptic Christianity in the sixth century, but there is no textual evidence to confirm whether the Meroitic numerical notation system existed during this late period. The Coptic and Ethiopic numerals, both of which are clearly derived from the Greek alphabetic numerals, were used widely in the region from the sixth century onwards.

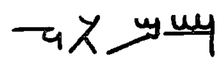
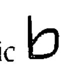
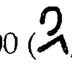
Egyptian Demotic

The demotic script developed in the late 8th century BC (Twenty-Fifth Dynasty), and began to replace the hieratic script about a century later. It was a cursive script consisting largely of consonantal characters, derived from the "business hand" used in the Nile Delta (Ritner 1996: 82). During the Late period and the Ptolemaic era, demotic writing was used very widely for administrative and literary purposes, and more sporadically throughout the Roman period. A set of ciphered-additive, base-10 numerals accompanied this script throughout its history.

The demotic numerals are probably the most neglected among numerical notation systems worldwide; Guitel (1975) and Ifrah (1998) note only their existence, while Menninger (1969) ignores them entirely. As with the hieratic numerals, there is a great deal of variation in the demotic numeral-signs; the ones presented in Table 2.18, from Sethe (1916: Table I), are typical of those found in papyri of the Late and Ptolemaic periods. Griffith (1909: 415-417) provides an interesting paleographic comparison of the demotic numeral-signs found on a selection of papyri dating from the Twenty-Sixth dynasty to the Roman period.

Table 2.18: Demotic numerals



The demotic numerals are a base-10, ciphered-additive system, written from right to left. Thus, 4637 would be written . They are less reliant on the cumulative principle than their hieratic ancestor (compare hieratic  and demotic  for 3). Some of the signs for the thousands may be vaguely multiplicative, as there is a general resemblance between the signs for the hundreds and the corresponding signs for the thousands. Nevertheless, it is more likely that they are simply further reductions of the non-multiplicative hieratic signs. Sethe (1916: Table I) suggests that additive phrases incorporating two lower signs (3000+2000, 4000+3000) were used for the missing 5000 and 7000 signs. Above 10,000, the demotic numerals, like the hieratic ones, are very clearly multiplicative (though such high expressions are fairly rare); for instance, Parker has found multiplicative expressions for 90,000 (, = 9 x 10,000?) and 100,000 (, = 10 x 10,000?) in his study of demotic mathematical papyri (Parker 1972: 86). As in the hieratic numerals, a small dot placed above a given integer indicated the corresponding unit fraction.

The demotic numerals are directly derived from the hieratic forms used in the 8th century BC. It is interesting that two systems that served similar functions and that had nearly identical structures should co-exist for such a long period. As the hieratic numerals were used as late as 200 AD, the two systems were used side by side in Egypt for nearly a millennium. This abnormality can be explained in part by regional variations, with Upper Egypt retaining the hieratic numerals and Lower Egypt using demotic. Unlike the corresponding writing systems, the hieratic and demotic numerals would have been largely mutually intelligible until the Ptolemaic period at least, which may have facilitated communication between different parts of Egypt. The demotic script and numerals were accorded royal preference in the Twenty-Sixth Dynasty, and thus they were used for most royal functions thereafter, while the hieratic system was retained primarily for calligraphic religious texts (Ritner 1996: 81-82).

Unlike the hieratic script and numerals, which were rarely written on stone except at the very end of their history, demotic inscriptions are found on stone as well as ceramics and papyrus. Like their predecessor, demotic numerals served a wide variety of commercial, legal, and other administrative functions, as well as indicating dates. A number of demotic mathematical papyri have survived from the Ptolemaic period, confirming the suitability of the system for arithmetical and mathematical purposes (Parker 1972; Gillings 1978). However, as with the hieratic numerals, this small selection pales in comparison with the enormous number of demotic texts that contain numerals but serve no mathematical function. Much of our paleographical knowledge of the demotic numerals comes from administrative texts, such as dowry records or educational papyri (Griffith 1909).

The importance of the demotic numerical notation system lies not in any structural feature or unusual function, but rather in its historical role as the immediate ancestor of several other numerical notation systems. The demotic numerals are almost certainly ancestral to the Greek alphabetic numerals (ch. 5). These numerals, which are structurally identical to the demotic numerals, first appear in the 6th century BC in Ionia and Caria, at which time Greek trade with Egypt was beginning in earnest, and the Ionian trading city of Naukratis in the Nile Delta was the major centre for trade between Egypt and Greece. Furthermore, the alphabetic numerals became common in the late 4th century BC, at which time Egypt came under Ptolemaic control. Remarkably, the similarities between the demotic and Greek alphabetic numerals have been substantially ignored over the past century, with most scholars inclined to treat the latter system as a case of independent invention (but cf. Boyer 1944: 159). Secondly, there are strong similarities between the demotic numerals and the Brāhmī numerals (ch. 6), which began to be used in India around 300 BC. In this case, the historical connection between the two regions is not as clear, but the structural similarities between the two systems suggest some connection. While trade between Egypt and India did not become common until the Roman period, there are strong indications of overseas trade dating from the Ptolemaic period and perhaps even somewhat earlier. Again, few historians of mathematics have proposed this connection, although it has held some popularity among Indologists for over a century (Bühler 1896, Salomon 1998).

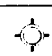
The demotic numerals continued to be used throughout the Ptolemaic era in various documents. By the Roman period, however, they were used increasingly rarely, as the general decline of Egyptian cultural institutions continued apace. However, even though Roman imperialism was the immediate circumstance surrounding the decline of the demotic numerals, they were not replaced with Roman numerals, but rather with the

Coptic numerals, which were themselves descended from the demotic through the Greek alphabetic numerals. As Christianity began to take hold in Egypt, and the Coptic script and numerals became more widespread, demotic suffered a fatal decline. The last demotic texts date to around 450 AD.

Linear A (Minoan)

The Linear A script was the standard script used in the Minoan civilization of Crete (and, to a lesser extent, other Aegean islands) between 1800 and 1450 BC (Bennett 1996: 132). It is perhaps the most famous of all undeciphered scripts, having foiled decades of effort to interpret it. Only the numerals and a few other ideograms for commodities can be deciphered. Linear A is written from left to right and is almost certainly a mixture of syllabograms and logograms. Its well-known numeral-signs are shown in Table 2.4 (Sarton 1936: 378; Ventris and Chadwick 1973: 36):

Table 2.4: Linear A Numerals

1	10	100	1000
			

The Linear A numerical notation system is decimal and cumulative-additive, and is written from left to right with the exponents in descending order. Where appropriate, signs are grouped in two rows of up to five signs each rather than listing them in an uninterrupted row. Thus, 7659 might be expressed as shown in Figure 2.5.

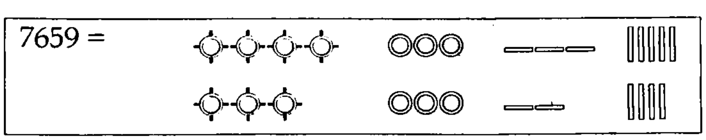


Figure 2.5: Linear A numeral-phrase for 7659

The variant dot symbol for 10 is found only in early Linear A documents and is probably related to the identical numeral-sign for 10 in the contemporaneous Cretan hieroglyphs. Other than this, however, the system remained unchanged throughout its history. While Evans (1935: 693) suggested that there may have been a sign X or + that stood for zero, this was later shown to be a sort of "check-mark" or sign for completion of an item, or perhaps served some other bookkeeping function (Bennett 1950: 205).

It is now thought that Linear A may be the earliest of the three Aegean numerical notation systems. It is probable that the Linear A script and numerals were borrowed from the Egyptian hieroglyphs, which have an identical structure (cf. Sarton 1936: 378). Trade between Egypt and Crete was extensive during the period when Linear A developed. Admittedly, there is no real similarity between the numeral-signs of the two scripts, except in the use of vertical strokes for the units, which is common to almost all systems used in the Mediterranean region. Whereas Egyptian hieroglyphic numerals are clearly pictorial and concrete representations, Linear A numerals are abstract and simplified. However, we would not expect the Egyptian signs (which have phonetic meanings in the hieroglyphic script) to be adopted by the Minoans, for whom the signs would have no such associations. The abstract and geometric character of the numeral-signs also makes it impossible to exclude an entirely independent origin for the system. Branigan (1969) has discussed just such a possible geometric precursor to the Linear A numerals, in which he speculates that concentric circles on sealings from Phaistos may have represented tens, hundreds, and thousands. Other derivations, however, such as the link suggested between Linear A and the Proto-Elamite numerals (ch. 7) of 4th millennium BC Iran, are dubious, as the chronological and geographical gaps are simply too large to be credible (Brice 1963). At present, the hypothesis of Egyptian origin remains the most likely explanation for the structure of Linear A numerals, with the form of the numeral-signs developed indigenously.

Numeral-signs are the only known means of representing numbers in Linear A; although it remains possible that lexical numerals were written using syllabic signs, the fact that the closely related (and deciphered) Linear B script does not do so suggests that this is unlikely. The vast majority of Linear A documents are clay tablets having an accounting or bookkeeping function, and thus we have many examples of the use of numerals. Vertical strokes that probably represented numbers have been found in other contexts, for example, on Minoan balance weights; these marks, however, do not show any clear relation to the Linear A signs found on the clay tablets and are probably simply unstructured unit-marks or tallies (Petruso 1978). Stieglitz has proposed that a Minoan numerical graffito found at Hagia Triada and containing the sequence of numbers (1, 1 1/2, 2 1/4, 3 3/8), in which each number is 1.5 times the previous one, represents a series of musical notes or tunings for a stringed instrument (Stieglitz 1978). While his theory is interesting, I think it more likely that the series served an economic function such as calculating interest. Since we do not have significant literary or monumental texts in Linear A, we do not know if the numerals were ever used in other contexts.

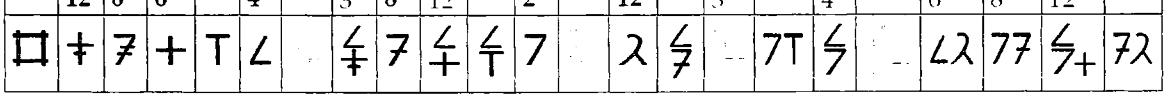
While the Cretan hieroglyphic numerals were formerly thought to be ancestral to Linear A, it now appears that Linear A predates the Cretan hieroglyphs, perhaps by as much as a century. The exact historical relationship between the two numerical notation systems is unclear, but I believe it most likely that the Cretan hieroglyphic numerals were a local variant of the Linear A system. The Linear B Mycenean script used on Crete and the Greek mainland is clearly derived from Linear A. Its numerals, discussed below, are nearly identical to those of Linear A, and are certainly derived from them. The precise relation between the peoples using the Linear A and B scripts is still unclear, as is the question of the cause of the collapse of the Minoan civilization in the 15th century BC. It is clear that the two scripts coexisted in Crete from about 1550 to 1450 BC. Presumably, during this period, the Greek-speaking Myceneans adapted Linear A for their own language, resulting in Linear B. No Linear A inscriptions are found after about 1450 BC, and its replacement by Linear B was complete throughout the Aegean world within about a century of the latter system's development.

Linear A Fractions

While it is beyond the scope of this work to discuss in detail systems for representing fractions, the base-24 system for fractional values used in Linear A is of some theoretical and historical interest. This system, though complex and not yet fully interpreted, has been studied extensively by Daniel Was, who has developed a logically consistent theory for its interpretation (Was 1971, 1974). The system comprises 23 signs for the fractions from 1/24 to 23/24, as shown in Table 2.5 (cf. Was 1971: 35-51; Struik 1982: 55).

Table 2.5: Linear A fractions

1/24	**2/24**	**3/24**	**4/24**	**5/24**	**6/24**	7/24	8/24	**9/24**	10/24	11/24	**12/24**	13/24	**14/24**	15/24	16/24	17/24	18/24	19/24	20/24	21/24	22/24	23/24
	1/12	**1/8**	**1/6**		**1/4**		1/3	**3/8**	5/12		**1/2**		**7/12**		2/3		3/4		5/6	7/8	11/12	



Nine of the signs are basic units (those listed in bold: 1/24, 1/12, 1/8, 1/6, 5/24, 1/4, 3/8, 1/2, 7/12) while the others are additive combinations of these signs. Four signs (those in grey: 7/24, 13/24, 2/3, 19/24) are unattested, and are hypothetical reconstructions. It seems odd that a number like 5/24 would have a basic sign while more common fractions such as 1/3 and 3/4 have composite ones, and that there is no sign for 2/3 at all. The values Was has assigned to these signs are not universally accepted (cf. Ventris and Chadwick 1973: 36 for an alternate interpretation). However, by using frequency-based methods and statistical analysis, coupled with trial-and-error techniques, Was has established these to be the most likely values for the various signs. This system demonstrates that the Western use of the same base for fractions as for integers is by no means universal or necessary. Rather, because there are certain metrological and commercial functions for which a base with many divisors may be useful for fractions (division into halves, thirds, and quarters), it may be the case that this specialized fractional system was developed to solve metrological problems and was not seen as part of the integral Linear A numerical notation system.

Struik (1982: 56) suggests that the system may have some relation to the Egyptian hieroglyphic "unit fractions". While this theory is attractive in light of my contention that the Linear A numerical notation system as a whole may have an Egyptian origin, there are several problems with this theory. While several of the basic Linear A fractional signs are unit fractions, three of them (5/24, 3/8, and 7/12) are not, and 1/3, which we would expect to be written as a unit fraction, is composite. The Egyptian hieroglyphs used a special sign for the non-unit fraction 2/3, but no sign for 2/3, basic or composite, has been found in Linear A. Finally, while the Egyptian unit fractions are derived from the corresponding integer (i.e., the sign for 1/24 is derived from the numeral-phrase for the integer 24), the Minoan fractions show no such resemblance to the appropriate integers. In short, I believe it is far more likely that the system is of indigenous origin, arising in the context of particular metrological problems. Regardless of its origins, the Linear A system was not transmitted to the related Mycenean Linear B numerals or to any other script, adding further weight to the hypothesis of its origin in a local metrological system.

Cretan Hieroglyphic

The Cretan Hieroglyphic or Pictographic script was first described by Sir Arthur Evans from his discoveries at Knossos, and is now generally thought to have arisen at about the same time as, or slightly later than, the Linear A script. Its use is generally thought to have lasted from 1750 to 1600 BC (Bennett 1996: 132). It is found on only a very limited set of seal-stones and clay impressions thereof. While the script is still undeciphered, it is probably of a mixed syllabic and logographic structure, like other Aegean scripts. Among the few signs able to be read in the Cretan hieroglyphs are its numerals, which are shown in Table 2.6 (Sarton 1936: 378; Ventris and Chadwick 1973: 30-31):

Table 2.6: Cretan hieroglyphic numerals

1	10	100	1000
/)	●	\	◊

The system is cumulative-additive and decimal, and usually written from left to right; thus, 3124 would be written as ◊◊◊\●●////. Since the Cretan hieroglyphs are largely undeciphered, it is difficult to speculate on the history of their numerals. As with other Aegean scripts, an Egyptian origin for the system has been proposed (Sarton 1936: 378), though this cannot be demonstrated conclusively. There is no great similarity between the numeral-signs for the Cretan hieroglyphs and any other system, except that the use of the dot for 10 is common with some early Linear A inscriptions. Nevertheless, unless we are to believe that the Cretan hieroglyphs and their base-10, cumulative-additive numerical notation system developed independently from both Linear A and the Egyptian hieroglyphs, some connection with the other systems in this family must exist. It is probable that the Cretan hieroglyphic numerals are a local variation of the Linear A numerals, or less plausibly, that they derived directly from the Egyptian hieroglyphic system. The contexts in which the numerals are found are similar to those for Linear A. The Cretan hieroglyphic inscriptions include information on commodities such as wheat, oil, and olives and thus are probably records of transactions, inventories of goods and similar administrative documents (Ventris and Chadwick 1973: 31). By around 1600 BC, Cretan hieroglyphs had been entirely replaced by Linear A.

Linear B (Mycenean)

The Linear B script was used on Crete and the Greek mainland in the middle to late 2nd millennium BC, and was used to write an archaic Greek dialect on clay administrative tablets. It is written from left to right, and consists of a syllabary with a large repertory of logograms and taxograms, including a numerical notation system. The Linear B numerals are shown in Table 2.7.

Table 2.7: Linear B numerals (Ventris and Chadwick 1973: 53)

1	10	100	1000	10000
𐄇	𐄐	𐄙	𐄢	𐄫

The Linear B signs are mostly identical with the Linear A signs, except that the sign for 10 is always a horizontal stroke (never a dot), and there is a sign for 10,000 that is not found in the earlier system. The 10,000 sign is probably a multiplicative combination of the signs for 10 and 1000. The structure of the system is cumulative-additive and decimal, with the highest exponents on the left, written in descending order and with five or more identical signs divided into two rows. Thus, 68357 might be expressed as shown in Figure 2.6.

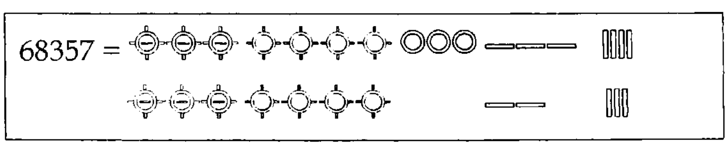


Figure 2.6: Linear B numeral-phrase

Unlike the Linear A numerals, Linear B lacks a separate system for expressing fractions; instead, specific logograms are used to express divisions of a given metrological unit and then combined with numeral-signs as appropriate (just as one might say 10 cm instead of 0.1 m). Ventris and Chadwick note, however, that some of the Mycenean logograms for metrological units are similar or identical to Minoan signs for fractions, and may have had their origins in specific ratios of two types of units, a point which further emphasises the indebtedness of Linear B to its Minoan forerunner (1973: 54-55).

The Linear B system definitely originated through direct contact with the Minoan civilization and the Linear A numerals. The earliest Linear B inscriptions date from the

16th century BC, so the two scripts coexisted on Crete for about a century. Their numerical notation systems are so similar that some authors do not distinguish between the two (Ventris and Chadwick 1973: 53; Struik 1982). I treat them as separate systems herein, recognizing that the distinction between the two is not nearly as great as between the two scripts, which record different languages. Throughout the history of the Linear B numerical notation, there is no observable change in the form of the numeral-signs or in the structure of the system.

Linear B numerals are found primarily on clay tablets serving accounting and financial purposes. Numerals are used both for counting discrete objects (men, chariots, etc.) as well as for measures of dry and liquid volume and weight. Almost all Linear B documents relate to administrative and bookkeeping functions, suggesting a very limited level of literacy and numeracy throughout Mycenean society. Even so, the consistency of the numerals throughout several centuries and across a substantial geographic area suggests that some sort of scribal education system was in place to transmit knowledge of both the Linear B script and its numerals. We do not know if Linear B numerals were written on papyrus or other materials, though such uses are certainly possible.

We also do not know whether the Myceneans used their numerals for arithmetical purposes. Anderson's (1958) theory on the means by which such calculations could be undertaken suffers from the defect that it involves aligning and manipulating numbers as one would in Western arithmetic, although there is no evidence that such a procedure was ever undertaken. Dow (1954: 32) and Anderson (1958: 368) both point to a clay tablet found at Pylos (designated Eq03) in which tallying in groups of five units is used to reach 137. Other tablets from Pylos discussed by Ventris and Chadwick (1973: 118-119) show that the Myceneans could successfully compute complex ratios in order to determine the contributions of goods required from towns of different size. Rather than proving that the Myceneans used numerical notation for arithmetic, however, these examples indicate that tallying by units and in groups of five, rather than the purely decimal-structured numerical notation, was the method used for computation. None of this denies that clay tablets recorded the results of rather complex computations done mentally, through tallying, or perhaps by some other method.

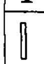
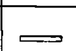
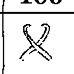
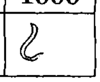
The question of possible descendants of the Linear B numerals is extremely interesting. There is no relationship between the Mycenean numerals and either of the numerical notation systems developed in archaic and Classical Greece (the acrophonic and alphabetic systems). It is conceivable, however, that there is some relationship between the Mycenean and Etruscan numerals (ch. 4). Both Haarmann (1996) and Keyser (1988) have raised this claim, which will be discussed in detail below when considering the origins of the Etruscan system. Mycenean settlements have been found in Sicily and southern Italy, providing one possible locus for cultural contact. However, this theory is controversial, not least because of the time elapsed between the latest known Linear B documents (12^{th} century BC) and the first Etruscan ones (7^{th} century BC). A more likely descendant of Linear B numerical notation is the Hittite hieroglyphic system, which was invented around 1400 BC and used by Hittite and Luwian speakers in Anatolia. The Hittite signs for 1 and 10 are identical to the Linear B ones, and when the Hittite numerals were developed, there were Mycenean settlements in western Anatolia (such as at Miletos) and on Cyprus that were engaged in trade throughout the eastern Mediterranean. The contemporaneity of the two systems makes this scenario plausible, if not proven.

The perplexing and apparently violent end of the Mycenean civilization in the 12^{th} century BC, and the repeated razing of major sites such as Mycenae and Pylos, marks the end of the Linear B inscriptions and the start of the "Dark Age" of Greek civilization. No writing or numerical notation of any kind is known from the Aegean region between 1100 BC and the introduction of the Greek alphabet a few centuries later.

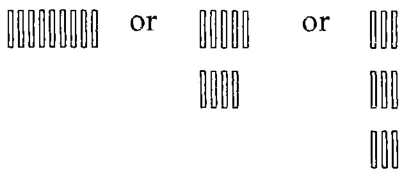
Hittite Hieroglyphic

The Hittites were an Indo-European population who lived in central Asia Minor from about the end of the 3rd millennium BC. The Hittite and closely related Luwian languages are the first Indo-European languages for which we have solid textual evidence. By the middle of the second millennium BC, two distinct scripts were in use in the Hittite Empire. Firstly, a cuneiform script (borrowed from Mesopotamia) was used to write the Hittite language. Its numerals are closely related to the Assyro-Babylonian cuneiform system, and so will be treated in Chapter 7. Additionally, an indigenous hieroglyphic script was used to represent the Luwian language on monumental inscriptions, on a few lead tablets, and probably also on wooden tablets that have not survived (Melchert 1996: 120). This script was used from about 1500 to 1200 BC, during the apogee of the classical Hittite Empire, and then is found only sporadically until the rise of the Neo-Hittite kingdoms between around 1000 and 700 BC, during which time it was again common (Hawkins 1986: 368). This script is known as Hieroglyphic Hittite or Hieroglyphic Luwian, and has a mixed syllabic and logographic structure. Among the purely ideographic signs, the Hittites used a set of written numerals as shown in Table 2.8 (cf. Laroche 1960: 380-400).

Table 2.8: Hittite numerals

1	10	100	1000
			

The system is purely cumulative-additive and uses a base of 10. Numeral-phrases were written from left to right, right to left, or top to bottom, depending on the overall direction of the inscription. Thus, the number 3635 might be written as . Like the Egyptian and Aegean systems, Hittite numeral-signs could be grouped in chunks or clusters of three to five unit-signs, but were also sometimes written in a single line; Laroche (1960: 395) indicates that 9 was variously written in the following three ways:



The most likely theory for the origin of the Hittite hieroglyphic numerals is that they were based on one of the Aegean numerical notation systems. Both the Linear A and Linear B scripts were in use around 1500 BC, when the first Hittite hieroglyphic inscriptions are found, but Linear A was almost extinct by that time. Like the hieroglyphs, the three Aegean scripts use a combination of syllabograms and logograms. However, because the Hittite syllabary is derived from the phonetic values of the ideograms in the Luwian language, it is very likely that the signs used in the script are indigenous to Anatolia. Considering only the issue of numerical notation, the Linear A, Linear B, and Hittite hieroglyphic systems are decimal and cumulative-additive, and all use a horizontal stroke for the units and a vertical stroke for the tens. There was a significant degree of intercultural contact between the Aegean and Asia Minor during this period. The Myceneans had settlements in western Anatolia and traded throughout the eastern Mediterranean, and were possibly the "Ahhijawa" (Achaeans) mentioned in the Hittite archive from Bogazkoy. Because the Luwian language was spoken primarily in western Asia Minor and only later came to be used in the Hittite Empire, the transmission of the numerals from the Aegean to western and then central Anatolia is plausible (Hawkins 1986: 374). An alternate hypothesis is that the Hittite system was based directly on the Egyptian hieroglyphic numerals, since the Hittites were in contact with Egypt at that time.

Due to the paucity of extant examples, little can be said about the function and use of the system. The numerals are found on a variety of stone inscriptions and lead tablets. Most notable among these are the Kululu lead strips (mid to late eighth century BC), which record village census data using an abundance of numerical signs (Hawkins 2000: 503-505). The Hittite numerical notation is used far more frequently than lexical numerals, which is also true of the Egyptian hieroglyphs and Aegean scripts. There is no discernable change in the structure or sign-forms of the system throughout its history, despite the fact that there is little evidence for its use between 1200 and 1000 BC, following the invasion of Phrygians and others who ended the classical Hittite kingdom. It is probable that during the next two centuries, the hieroglyphs were used only on perishable materials, such as wooden tablets, although no direct evidence of this is available (Hawkins 1986: 374).

The only certain descendant of the Hittite hieroglyphs is the unstructured Urartian system, which I discuss briefly below, but this system does not appear to have been a full-fledged numerical notation system. Another possible descendant are the numerals that accompany the Cypriote syllabary, which was invented around 800 BC. The proximity of the neo-Hittite kingdoms to Cyprus, the extensive trade relations between the regions, and the identical structure of the two systems, all suggest that such a derivation is likely. However, there are too few inscriptions in the syllabary that contain numerals to establish an accurate chronology or even to secure values for certain numeral-signs. Other possible descendants of the Hittite hieroglyphic system are the earliest Levantine systems, Phoenician and Aramaic (ch. 3). However, these systems developed around 750 BC, at the very end of the Hittite system's history, and are structurally distinct from it, since they have a sign for 20 and are multiplicative-additive above 100.

After a hiatus of about 200 years, the Neo-Hittite kingdoms resumed using the hieroglyphic numerals on monumental inscriptions around 1000 BC. However, their eventual subjugation to the Assyrian empire ended this usage by around 700 BC, and it was replaced for all functions by the Assyro-Babylonian common numerals. Later scripts and numerical notation systems developed for related languages of Asia Minor, such as

Lycian, were based on a Greek model and display no obvious relation to the Hittite hieroglyphs.

Urartian

A few inscriptions on clay jars found at the Urartian site of Altintepe (in eastern Asia Minor) use a syllabary closely related to the Hittite hieroglyphs to write single words in the Urartian language, starting in the early 8^{th} century BC (Laroche 1971, Klein 1974). In addition to these words, many of these inscriptions contain numeral-signs for small numbers using either 'pitted' dots or vertical strokes to represent units (i.e. 5 = ●●●●● or |||||). However, this system was used for only a handful of inscriptions and was never used for numbers greater than eight, making this system an unstructured tally-system having no numerical base. Klein's assertion that this usage "should thus be viewed as an isolated and short-lived phenomenon, possibly not outlasting the career of a single (foreign?) scribe" seems entirely warranted (1974: 93). Accordingly, Ifrah's inclusion of this Urartian system as a distinct numerical notation system seems unwarranted (Ifrah 1985: 139). For later (late 8^{th} century BC onward) Urartian inscriptions, this hieroglyphic script was supplanted by a cuneiform script.

Cypriote syllabary

As its name would suggest, the Cypriote syllabary was a purely syllabic script used only on the island of Cyprus. It was used between about 800 and 200 BC for writing the Greek language, and thus co-existed with the much more prominent and long-lasting Greek alphabetic script (Bennett 1996: 130). Cypriote is always written from right to left. None of the synthetic works concerning numerical notation have dealt with the (admittedly small) evidence for a distinct Cypriote numerical notation system, including Ifrah (1998), whose coverage of obscure systems is generally thorough. However, Masson (1983: 80), whose discussion of the Cypriote syllabary is the most detailed presently available, presents about a dozen inscriptions in which the system shown in Table 2.9 was used.

Table 2.9: Cypriote numerals

1	10
ǀ	—

This rudimentary system was decimal and cumulative-additive and, like the syllabary itself, was written from right to left. The numbers expressed using the system are very low; unless certain undeciphered signs are in fact numeral-signs (as discussed below), the highest number expressed in any Cypriote inscription is 22. This system obviously parallels the Aegean Linear systems from which the Cypriote numerals are probably derived. This is strongly suggested by the use of the Cypro-Minoan script on Cyprus as early as 1500 BC, which was very probably borrowed from Linear A. This simple derivation is made more complex, however, by the fact that eastern Cyprus was under Phoenician domination well into the period of the use of the syllabary. Masson rightly points out that the Phoenician numerical notation system is also written from right to left, uses vertical strokes for units and horizontal strokes for tens[4]. Furthermore, Masson notes the use of two unusual symbols: ◎, found in but a single inscription but possibly indicating 100 on the model of the Aegean systems, and Y, also in only a single document, but possibly signifying 20 (Masson 1983: 80). It is notable that the Phoenician system used O and H at various times as the sign for 20. Because Cypriote inscriptions do not contain dates, it is often difficult to place them in chronological context, but it seems possible either that the Cypriote system was borrowed to create the Phoenician one, or vice versa. A final complexity is that the Hittite hieroglyphic numerals, which

[4] In fact, the Phoenician 10-sign normally has a tail (—), but the analogy seems significant nonetheless.

were still in use in the Neo-Hittite kingdoms in 800 BC, also use a vertical stroke for 1 and a horizontal one for 10. Trade between Cyprus and the Hittites was substantial and it would have been an extremely short sea voyage between the two regions. None of this material categorically excludes the possibility that the aberrant signs found by Masson are non-numerical and that the Phoenician, Hittite, and Cypriote numerals are unconnected except by their temporal and geographic proximity on the island of Cyprus. The corpus of inscriptions containing numerical signs is simply too limited, and the numbers expressed too small, to resolve the issue of their origin.

Summary

Despite the enormous amount of work being done in the archaeology of the eastern Mediterranean, the genetic relations among the systems of this family have not been analyzed adequately in the past. The connections between the Egyptian hieroglyphic, hieratic, and demotic systems are well-established, but more data are needed to establish the specific links between the Egyptian and Aegean systems. Nevertheless, on the basis of a shared set of features that distinguish it from other, superficially similar families such as the Levantine (ch. 3) and Italic (ch. 4), the inclusion of all the hieroglyphic systems in a single family is warranted. First, all the hieroglyphic systems have a base of 10, but do not use a sub-base of 5 or additional structuring signs. Second, they mostly have a cumulative-additive structure, although the hieratic, demotic and Meroitic systems are ciphered-additive reductions of the original cumulative structure of the family. Third, large numbers of cumulative signs in a numeral-phrase are grouped in sets of three to five. Fourth, their direction of writing can be quite variable (left-right, right-left, top-bottom, or boustrophedon). Finally, unlike other systems in use in the Mediterranean, the hieroglyphic numerical notation systems are used far more frequently than the full phonetic writing of lexical numerals.

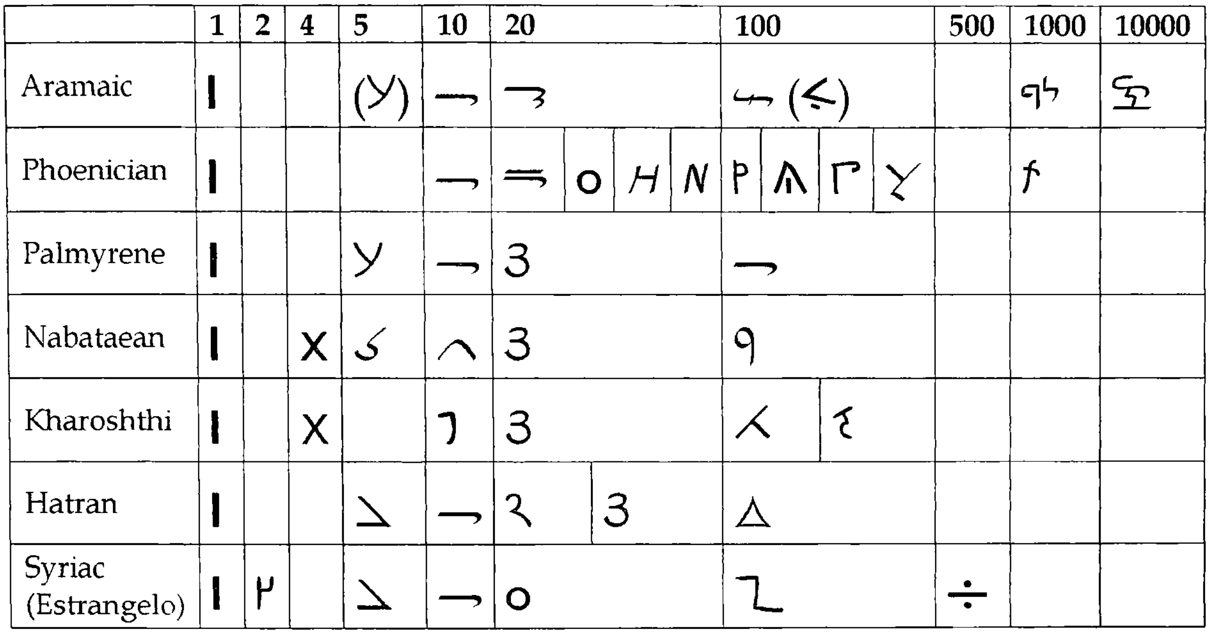
While no members of the Hieroglyphic family discussed in this chapter survived past 400 AD, its less direct descendants include the Roman numerals and probably even our own Western numerals (though greatly transformed). In the following four chapters, I will discuss a) the Levantine family (ch. 3), the Phoenician-Aramaic numerals and related systems; b) the Italic family (ch. 4), the Etruscan and Roman numerals and their descendants; c) the Alphabetic family (ch. 5), the Greek alphabetic numerals and related systems; and d) the South Asian numerals (ch. 6), the Brāhmī system and its descendants. While they are distinct enough to warrant placing them in separate families, all originate directly from the systems of the Hieroglyphic family.

Chapter 3: Levantine Systems

The first millennium BC was an era of considerable interregional commerce, warfare, and colonisation in the Levant. Living in a peripheral region of both Egypt and Assyria, Levantine peoples, such as the Aramaeans and Phoenicians, were exposed to a variety of cultural influences. The various Levantine numerical notation systems developed in the first millennium BC share a number of common features that reflect their debt to both Mesopotamia and Egypt, while allowing for considerable inventive energy among the indigenous creators of the systems. While this family of numerical notation systems was developed and most widely used in the Levant, it would eventually find its way into Asia Minor, Arabia, Persia, and India. The Aramaic notation is the most important and long-lived of the Levantine family, which also includes the Phoenician, Palmyrene, Nabataean, Kharoshthi, Hatran, and Syriac Estrangelo systems.

Unfortunately, despite their widespread use over a large geographical area, these systems have been ignored by scholars of numeration. Guitel (1975: 200) dismisses them as too irregular and primitive to be of real interest. Ifrah (1998: 227-234) devotes some detailed attention to the Levantine systems and correctly points out the common ancestry of these systems, but he fails to discuss the Kharoshthi numerals at all and attributes an Assyro-Babylonian ancestry to the entire group, which I will question below. For most purposes, we must turn to the earlier work of epigraphers and paleographers such as Schroder (1869), Duval (1881), Lidzbarski (1898), Cooke (1903), and Cantineau (1930, 1935) for analysing Levantine numerical notation. Despite the age of these works, there seems no reason to question the data presented. However, this tradition of scholarship was primarily oriented towards the study of the literatures of specific societies; thus, while the structure of the system and numeral-signs are well understood for each specific Levantine society, questions of diffusion and cross-cultural comparison remain largely unanswered. The numerical notation systems of the Levantine family (including the most common variants of the numeral-signs) are shown in Table 3.1.

Table 3.1: Levantine numerical notation systems



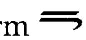
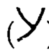
Aramaic

The Aramaeans, who originally inhabited a large portion of modern-day Syria, are first recognisable in the archaeological and written records around the end of the second millennium BC. During the ninth and eighth centuries, Aramaeans ruled a number of small states in the Levant, until these came under the domination of the Assyrian empire. Around this time, they developed a consonantal script on the model of the pre-existing Phoenician consonantary. Later, when Aramaic became the *lingua franca* of the Achaemenid Persians, this script diffused quite widely throughout the Levant, the Middle East, and parts of India. While the Aramaic numerical notation system that developed around the same time never achieved such an exalted status, it provides an interesting comparison of the relative influences of Egyptian and Mesopotamian civilisations on the Levant in the mid-to-late first millennium.

The earliest Aramaic numerical notation system (used between the 8th and 3rd centuries BC) had distinct signs for 1, 10, 20, 100, and 1000. While it has a distinct sign for 20, it is largely a decimal system, not a vigesimal one. The signs of this system are shown in Table 3.2.

Table 3.2: Aramaic numerals

1	10	20	100	1000
𐡘	𐡛	𐡜	𐡝	𐡞

The system is a purely cumulative-additive one for numbers up to 99, written (as with the script itself) from right to left, using signs for 20, 10, and 1. The unit-signs are grouped in threes (as in the systems of the Hieroglyphic family), since up to nine such signs could be required. Occasionally, when an ungrouped unit-stroke was present in a numeral-phrase, it was written at a slight angle. Because there was a distinct sign for 100, no more than four 20-signs and one 10-sign would ever be required, obviating the need for such groupings for higher values. The ten-sign appears to have originally been a simple horizontal stroke, with a tail added cursively. The 20-sign is almost certainly a ligatured combination of two ten-signs, particularly considering that a variant form  was sometimes used. There is a gradual trend throughout time towards the use of a special sign for 5 (), which Lidzbarski (1898: 199) notes appearing on an Assyrian brick as early as 680 BC. However, the majority of Aramaic numeral-phrases do not use the symbol for 5.

Above 100, the Aramaic numerical notation system is multiplicative-additive rather than cumulative-additive, and it is thus a hybrid system. To form 800, for instance, eight unit-signs (appropriately grouped) were placed in front of the sign for 100 in order to indicate that the values should be multiplied. The same principle was followed for the thousands. There were apparently two signs for 1000; the first, 𐡞, is actually no more than an abbreviated form of the Aramaic lexical numeral 'thousand' (Gandz 1933: 69-70), while the second, , is the same as the corresponding Phoenician numeral-sign (Lidzbarski 1898: 201-202). Thus, 2894 might be written as depicted in Figure 3.1.

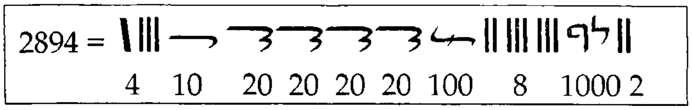


Figure 3.1: Aramaic numeral-phrase for 2894

While there is no distinct sign for 10,000 in the Aramaic system used in the Levant (though see below for Egyptian variants), numbers greater than 9,999 could be expressed using 10- and 20-signs in conjunction with the sign for 1000. Such numbers are rarely attested, however, and fractions were normally written out in words. Fractions are apparently found in a handful of inscriptions in which ungrouped unit-strokes |||| and ||||| are used to mean 1/4 and 1/5, and one inscription contains a special sign for 2/3 () (Lidzbarski 1898: 202).

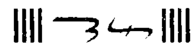
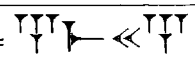
The origins of the Aramaic system are somewhat uncertain. The first Aramaic inscription with numerical notation is an 8th century BC ostracon from Tell Qasile, in which 30 is expressed as three horizontal strokes () (Lemaire 1977: 280). This is obviously not in the form expressed above. However, it may be a Hittite hieroglyphic numeral-phrase (ch. 2), since that system was still in use in the 8th century BC in the neo-Hittite kingdoms to the north. The first uncontestable example is found on an Assyrian bronze lion-weight found at Nineveh which dates to the late 8th century BC, on which 15 is expressed in three different ways on its three lines of text: in Aramaic number-words, as fifteen ungrouped single strokes, and according to the structure detailed above () (Cooke 1903:192). This threefold repetition using different methods of representation suggests that the system was unfamiliar at that time and place, either because it was new or because of its Assyrian context. A mid-eighth century BC date of origin of the Aramaic numerical notation thus seems reasonable.

The question of where and when its development took place is almost entirely unknown. While the Aramaic script almost certainly developed from the earlier Phoenician, there is no evidence that the Phoenicians used numerical notation before the 8th century BC. Chronologically, then, the appearance of the Aramaic and Phoenician systems is virtually simultaneous, making it difficult to establish which (if either) was ancestral to the other. Since the two systems were initially very similar, this question of priority is inconsequential.

Some scholars have suggested a Babylonian origin for the Aramaic system, given that the Aramaic numerals are found in Babylonian contexts early in their history. Furthermore, there are structural similarities between the Aramaic system and the Assyro-Babylonian common system (ch. 7), with which it shares a decimal base and the use of multiplicative-additive structuring for the hundreds and thousands (Gandz 1933: 69, Ifrah 1985: 356). There is much to be said for this argument, inasmuch as the structure of both systems is cumulative-additive up to 100, but multiplicative-additive thereafter. Furthermore, the presence of the earliest Aramaic numeral-signs in Assyrian contexts, as described above, supports this conclusion. The conquest of the Aramaeans in 732 BC by the Assyrian empire establishes a clear historical context in which this transmission could have taken place.

Yet an element of ambiguity in this simple derivation is that the Aramaic system is also similar to the Egyptian hieroglyphic system. Aramaic-speakers would certainly have had considerable contact with Egypt in the 8th century BC, and by the 6th century BC the Aramaic script was being used by settlers in Egypt at Elephantine and Saqqara. There are a number of similarities in the forms for signs. Like the Egyptian hieroglyphic but unlike the Assyro-Babylonian common system, Aramaic uses vertical unit-strokes grouped in threes to express the units. A relationship between Aramaic 𐡛 and hieroglyphic 𓎆 (both signifying 10) has also been postulated (Schroder 1869: 186). This may be overstating the case, since the hooked Aramaic sign may simply be a cursive alteration of a horizontal stroke. Regardless, both are very different from the cuneiform Assyro-Babylonian system. The Aramaic use of unit fractions along the Egyptian hieroglyphic model, including the exception of having a special sign for 2/3, further suggests Egyptian borrowing. Finally, Egyptian hieroglyphic numeral-phrases are primarily written from right to left, as in Aramaic, whereas the Assyro-Babylonian system runs in a left-right direction. These differences are compared in Table 3.3.

Table 3.3: Aramaic, Egyptian hieroglyphic, and Assyro-Babylonian numerals

Aramaic	424 = 
	4 20 100 4
Hieroglyphic	424 = 
	4 20 400
Assyro-Babylonian	424 = 
	4 100 20 4

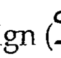
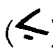
To muddy the waters even further, two other systems of the Hieroglyphic family were in use in the eastern Mediterranean around 750 BC and could potentially have been ancestral to the Aramaic numerals in place of the Egyptian system. The neo-Hittite kingdoms, although on the wane by that time, were still present in southeastern Anatolia, immediately abutting the Aramaeans. Moreover, the Cypriote numerals were invented just before that time, and there was enormous trade between Cyprus and the Levantine coast. Both of these systems are cumulative-additive and decimal and use vertical strokes for 1 and horizontal strokes for 10, and thus are equally plausible ancestors of the Aramaic system as the Egyptian hieroglyphic numerals.

I believe that the Aramaic numerical notation was developed under dual diffusion from Egypt and Mesopotamia, and for this reason is best placed at the head of its own family. Its direction of writing and certain numeral-signs are similar to Egyptian, while its structure is very similar to the Assyro-Babylonian common system. Geographically and historically, the Aramaeans and other Levantine peoples were peripheral to both civilisations in the mid-first millennium BC, at the time of the system's invention. I further believe that the failure to recognize this dual debt can be explained in terms of the way in which phylogenies are constructed for cultural phenomena. As cultural phylogenies for scripts and numerical notation systems are almost always arranged in accordance with a biological taxonomic scheme, it is rarely recognized that a phenomenon may have multiple origins, each making a contribution to the descendant, much as biological parents contribute to a child's genetic makeup. One would not wish to draw the analogy too strictly, of course, but we should nevertheless recognise that phylogenies for sociocultural phenomena may have quite different structures than biological ones.

If this explanation is true, we need to ask why the Egyptian hieroglyphic numerals, rather than the hieratic, would be chosen as a model for the Aramaic numerals. As I discussed in Chapter 2, the hieratic numerals were widely used among the Hebrews in the early part of the 1st millennium BC. It is highly unlikely that the Aramaeans were completely unfamiliar with the hieratic numerals. However, the dissimilarity of the Aramaic and hieratic numerals excludes this possibility. Like Millard (1995: 190-91), I find the failure of the Aramaeans to adopt the hieratic numerals to be rather curious. Other hypotheses, such as the derivation from Hittite or Cypriote numerals rather than the Egyptian hieroglyphs, are plausible, and we may not have a good answer to this question unless more data are forthcoming.

Of course, the existence of a distinct sign for 20 in Aramaic, and the fact that it recombines features of two quite different systems, suggests that we should attribute a good deal of inventive energy to the Aramaeans themselves. As Ifrah (1998: 136) points out, in the Semitic family of languages, the word for 'twenty' is etymologically the dual of 'ten'. This may explain why the 'etymology' of the Aramaic numeral-sign for 20 is that of two ligatured 10-signs. This development of a special sign for 20 outside the regular decimal base of the numerical notation system appears to be a unique development of the

Levantine numerical notation systems; neither the Egyptian hieroglyphic system nor the Assyro-Babylonian systems have this feature.

Like the script to which it was attached, the Aramaic numerical notation spread very widely throughout the second half of the first millennium BC. Segal (1983) gives ample evidence for the use of the system among Aramaic texts from Saqqara in Lower Egypt throughout the 5th and 4th centuries BC, and Aramaic papyri found at the 5th century BC military colony at Elephantine show the use of the system in numerous administrative documents. While the system as used in the Levant had no special sign for 10,000, the Aramaic papyri found at Saqqara and Elephantine do use such a sign (), which obeys the multiplicative principle in the same way as detailed above for 100 and 1000 (Segal 1983: 131; Ifrah 1985: 335). An alternate sign for 100 () was also used in Egyptian Aramaic, but it resembles none of the signs used in the Levant and is not similar to any of the Egyptian demotic or hieratic signs used at that time.

The Aramaic script was widely used throughout the Achaemenid Empire from the 6th to 4th centuries BC on clay administrative and legal tablets and in inscriptions on stone monuments. While the scripts used in official royal proclamations and dedications were Old Persian, Babylonian, and Elamite, Aramaic was the *lingua franca* of the Empire and was used for most administrative functions. As such, it was used as widely as Lower Egypt, Asia Minor, and the Transcaucasus and even as far east as the Indus River. Throughout its history, the Aramaic numerical notation was used extensively on monumental inscriptions, ostraca, and administrative papyri. However, in literary and ritual contexts, numbers were written using lexical numerals only. In fact, numerical notation systems of the Levant as a whole tend to be used only sporadically, with lexical numerals sufficing for most purposes. There is no evidence that Aramaic numerals were ever used for mathematics or even for doing arithmetical calculation, although they were used to record the results of calculations used in commerce and administration. Any mathematics conducted by users of the Aramaic script would have been written with the

Babylonian sexagesimal positional system (ch. 7), the Egyptian demotic numerals (ch. 2), or the Greek alphabetic numerals (ch. 5).

The end of the Achaemenid Empire did not spell the end of Aramaic influence over the Middle East; however, it did result in the fragmentation of what previously had been a unified script and numerical notation into several regional variants. After the Alexandrine conquest, Aramaic inscriptions become somewhat less common for a century or so under the Hellenistic Seleucid kingdom. During this period, Greek alphabetic numerals were often used administratively. Only in the mid-to-late 2nd century BC are Aramaic inscriptions found again with great frequency. By this time, political and ethnic divisions in the Levant and Persia had led to variant numerical notation systems. The Hellenised Palmyrene, Nabataean, Hatran, and Edessan Syrian populations of the Levant each possessed their own variant numerical notations based on Aramaic. In these variants, the use of a distinct sign for 5 was far more prominent than in Aramaic numerals. The Kharoshthi numerical notation used in the Hindu Kush, which reached its mature form in the early 1st century BC, is also clearly a variant form of Aramaic. Each of these will be discussed briefly below. It is unlikely that the Brāhmī numerical notation (ch. 6), which is structurally quite different from the Levantine systems, was derived from Aramaic, even though the Brāhmī script is of Aramaic ancestry.

Phoenician

The Phoenicians, who inhabited various cities along the Levantine coast in the first millennium BC (Tyre and Sidon being foremost among them), were perhaps the greatest mercantile people of the ancient Mediterranean. While their consonantal script was developed late in the second millennium on the model of the earlier Canaanite consonantary, the Phoenicians did not possess a distinct numerical notation system until several centuries later.

The Phoenician numerical notation system is similar in structure to the Aramaic, with distinct signs for 1, 10, 20, 100, and 1000. These signs (including some paleographic variants) are shown in Table 3.4.

Table 3.4: Phoenician numerals

1	10	20				100				1000
I	¬	=	o	H	N	P	Λ	Γ	Y	f

Like Aramaic, this system is purely decimal with the exception of the 20-sign, cumulative-additive below 100 and multiplicative-additive thereafter. Unit-signs are simple vertical strokes, although a left-slanting stroke is often used for ungrouped single strokes, and are grouped in threes, as in the Egyptian hieroglyphic and Aramaic systems. Like the Phoenician script itself, Phoenician numeral-phrases are always read from right to left. The most notable feature of Phoenician notations is the wide variety of forms for number-signs, particularly for 20 and 100. Schroder (1869: 188-9 and Table C) lists over 20 variants each for these two numbers, some of which can be attributed to differing paleographic styles, while others may reflect regional or diachronic variation. I will list only the more common forms for the sake of brevity. The 1000-sign is extremely rare; Schroder (1869) and Gandz (1933) do not report its existence, while Lidzbarski (1898: 201) reports only a single instance from Tyre. There is no evidence whatsoever for the use of a distinct sign for 5, in contrast to many Levantine systems, nor is there any evidence of numeral-signs for fractions.

For numbers greater than 100, a multiplicative-additive structure was employed as in Aramaic; a group of cumulative unit-signs preceding a single 100-sign indicates multiples from 100 to 900, with any additional signs to the left indicating the component of the number less than 100. Thus, 677 would be written as \||| ||| ¬HHH Λ||| |||. It is likely that the rare sign for 1000 also combined multiplicatively with sets of grouped unit-signs.

It is sometimes claimed that the Phoenicians used an alphabetic (presumably ciphered-additive) numerical notation system as early as 900 BC (Dantzig 1954: 295; Zabilka 1968: 117-119). The myth of Phoenician alphabetic numeration has been repeated for more than a century, but there is no foundation for this assertion. The first alphabetic numerals were developed by the Greeks in the late 6th century BC (ch. 5). Zabilka (1968: 118) claims that the first ten letters of the Phoenician alphabet were used on coins minted at Sidon to represent the numbers 1 through 10, based on Harris (1936:19), who is, however, referring only to Alexandrine coins. By this period in time, the Greek alphabetic numeral system was used throughout the Levant by speakers of both Indo-European and Semitic languages. Even this does not prove the existence of a true alphabetic numerical notation system among the Phoenicians in the 4th century BC; it could rather indicate a system of letter-labelling as used by the Greeks (Tod 1979), which is not really different from modern writers who label points of discussion A, B, C, and so on.

The first example of numerical notation in a Phoenician inscription is the Karatepe inscription of around 750 BC, which contains a single stroke for 1 (Millard 1995: 191). If this is a true example of the above system, then its appearance is virtually simultaneous with that of the Aramaic system. However, one unit-stroke is scant evidence for this. Like Aramaic, the Phoenician system is very likely modelled on both the Egyptian and Assyro-Babylonian common system, though whether Phoenician or Aramaic was developed first is not answerable at present.[1] Again, contact with the neo-Hittites or with Cyprus may also have played a role in the origin of this system. It was used on stone inscriptions, ink writings on clay, to a certain extent in administrative documents, and at

[1] Schroder (1869: 186-189) argues for diffusion from Egypt alone, but ignores the use of the multiplicative-additive structuring in the hundreds position in so doing.

a somewhat later period on coins. It was often used to enumerate regnal years and for record keeping of quantities of commodities. However, it was never used directly for arithmetical computation.

The Phoenician system was used from roughly 750 to 100 BC. However, during this time, Phoenicia was politically dominated in turn by the Assyrian, Neo-Babylonian, Achaemenid Persian, and Alexandrine Greek empires. The Phoenician numerical notation thus predominated in the Levant only during its early history. However, in the Phoenician colonies in North Africa and Spain (including, most importantly, Carthage), Phoenician and its Neo-Punic variant script continued to make use of the system detailed above, without significant regional variation, until Roman and Greek conquests in the 2nd century BC effectively ended its use. Coins from Akko, Tyre, and Sidon used Greek alphabetic numerals as early as 265 BC, though at Arvad and Marathus, Phoenician numerals were used on coins until about 110 BC (Millard 1995: 193).

The fruitful transmission of the Phoenician consonantal script throughout the Aegean and the Middle East has led some to speculate as to the transmission of its numerical notation system. However, the Levantine systems seem more likely to have developed from an Aramaic ancestor, and are generally regarded as such by scholars. Millard argues that Phoenician may have been the model for the Greek acrophonic (base-10, sub-base 5, cumulative-additive) numerical notation system (Millard 1995: 192). However, the acrophonic system's sub-base of 5, coupled with the more obvious derivation of acrophonic numerals from the very similar Etruscan system, makes such an origin unlikely. Similarly, while it has sometimes been argued that the Berber numerical notation system (ch. 4) of North Africa is of Phoenician origin, I will argue later that this development is later and part of the Italic family of systems. Finally, Schroder (1869: 187f) suggests that the Lycian numerical notation is a variant of Phoenician; again, however, its numeral-forms much more closely resemble the Greek and Roman than the Phoenician.

Palmyrene

Palmyra was an important mercantile city located in modern Syria whose inhabitants, Aramaic-speaking Semites, managed to retain considerable control over their own affairs despite Greek and Roman influence in the area. Palmyrene inscriptions are found dating from the 1st century BC to the mid-3rd century AD, continuing the tradition of the earlier Aramaic script. Palmyrene numerical notation retained much of the structure of the older Aramaic system, while introducing new numeral-signs. Along with the Nabataean, Hatran, and Estrangelo systems, Palmyrene numerical notation represents the final stage in the evolution of the Aramaic system. The Palmyrene system had distinct signs for 1, 5, 10, and 20, as shown in Table 3.5.

Table 3.5: Palmyrene numerals

1	5	10	20
𐡹	𐡽	𐡾	𐡿

These four symbols served to express any number less than 100. While in earlier Aramaic scripts the sign for 5 appeared only sporadically, it was a fundamental part of the Palmyrene system. Because of this, only four unit-signs were required at most; thus, the practice of grouping sets of unit-signs into threes was not needed. Like its Aramaic ancestor, Palmyrene numerical notation is strictly decimal and cumulative-additive below 100. For numbers greater than 100, Palmyrene, like Aramaic, is multiplicative-additive, but a new feature is introduced: the sign for 100 is identical to that for 10. The possibility of confusion is avoided by the requirement of having one or more unit-signs before the 100-sign, whereas no such signs could precede a 10-sign. Thus, the number 178 would be expressed as shown in Figure 3.2.

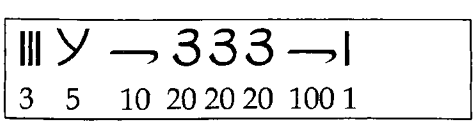


Figure 3.2: Palmyrene numeral-phrase

While this feature more clearly resembles the use of the positional principle than is found in the other Levantine systems, such phrases are multiplicative, not positional. To represent 100, the sign had to be combined with unit-signs; alone, it always meant 10, not 100. Cantineau (1935: 36) contends that the original Palmyrene sign for 100 was a horizontal stroke placed above a dot, but that it was later reduced until it was identical with the 10-sign. If so, the identity of the two signs may be largely coincidental.

Palmyrene numerical notation was restricted geographically and temporally to the city of Palmyra during the period from about 100 BC to 275 AD. During that time, however, it was used quite widely on inscriptions and records of commercial transactions, though not normally in religious or literary contexts. The importance of Palmyra as a Roman commercial centre rested on its strategic location and trade ties with peoples outside the Empire. There is no evidence for the use of its numerical notation for arithmetical or mathematical purposes aside from recording the results of calculations; the means by which these computations were performed is still unknown.

Despite considerable Hellenisation and Latinisation, Palmyra retained its script and numerical notation through the 3rd century AD, though Greek alphabetic and Roman numerals became more frequent for administrative and mercantile purposes. In 273 AD, following the short-lived independent rule of Queen Zenobia over the province (266-272 AD), Palmyra was destroyed by the Roman emperor Aurelian, abruptly ending its importance as a commercial centre. It is thus evident that political factors, rather than criteria of function and efficiency, led to the complete replacement of the Palmyrene numerical notation system by those of Greek and Roman colonisers. It has sometimes been argued that Palmyrene is ancestral to the Syrian Estrangelo numerical notation, though I will show below that this is only one of many possible scenarios of transmission.

Nabataean

The Nabataeans were a South Semitic people of Arabian ancestry who inhabited the area between Syria and Arabia in the southeastern Levant in the late 1st millennium BC and into the Christian era. Though not Aramaeans, they came under considerable Aramaean influence and adapted the Aramaic script for their South Semitic language, including a variant of its numerical notation. This system was used from approximately 100 BC to 250 AD in inland areas of the Levant (modern southern Syria and Jordan) including the cities of Damascus and Petra, and even as far south as the port of Aqaba on the Red Sea. Its signs are indicated in Table 3.6.

Table 3.6: Nabataean numerals

1	4	5	10	20	100
I	X	𐢫	𐢬	𐢭	𐢮

As with all the Levantine systems, Nabataean is decimal, cumulative-additive below 100 and multiplicative-additive above, with additional signs for 4, 5, and 20. The sign for 1 is obviously common to all the Levantine systems. Unit-strokes are grouped in threes where necessary and are sometimes joined together at the base in cursive writing. The sign for 4 is only used in some inscriptions, and then only in numeral-phrases for 4; 8 is expressed as III𐢫 (5+3) and II III III, but never to my knowledge as XX. Lidzbarski (1898: 199) argues that its shape represents four unit-strokes placed in a cross, strictly on graphic principles, but this is unproven. Its historical connection with the identical Kharoshthi sign for 4 is still unclear, but some link seems probable, given that the Nabataeans were frequently engaged in commerce with peoples to the east. On the other hand, Gibson (1971: 13) notes that the 8th century BC Samaria ostraca, in which the Hebrew variant of the Egyptian hieratic numerals (ch. 2) predominate, use a "+" or "x" shaped sign for 4, and this of course would antedate either the Nabataean or Kharoshthi symbol by several centuries. Finally, Cantineau (1930: 36) and Lidzbarski (1898: 199) believe the signs for 4 and 5 to be quite late inventions, possibly independent from any other system. More evidence is needed before a definitive answer can be given to this question.

The Nabataean sign for 10 is a more arched version of the hooked horizontal stroke used in most Levantine systems, while the 20-sign can be shown easily to derive from the Aramaic form. The sign for 100 is not obviously related to that of any other notation, though Cantineau (1930:36) argues for its possible derivation from Phoenician P. The sign for 100 combines with signs for 1, 4, and 5 multiplicatively Accordingly, the 4-sign is used to express 400, as in an inscription from Dumêr (near Damascus) from 94 AD in which the number 405 is expressed as ʃ9X (4 x 100 + 5) (Cooke 1903: 249). No Nabataean writings contain numbers higher than 1000.

The Nabataean numerical notation system is found on inscriptions dating from around 100 BC to the late 4th century AD, primarily in the inland Levant from Damascus south to Petra. Throughout its history, it was used in inscriptions on edifices, on ostraca, and on coins, but, like all Levantine systems, was apparently never used for mathematics or arithmetic. Though the Nabataeans were politically subordinate to Rome throughout most of the period under consideration, they held a monopoly over the caravan trade that passed from inland Arabia to the Levantine coast. Nabataean numerical notation has been found on economic documents and inscriptions in Greece, Italy, and Egypt.

Nabataean numerical notation continued to be used regularly until the 3rd century AD, at which time it began to be replaced by the Greek alphabetic and Roman numerical notation systems. The Nabataean script is ancestral to the neo-Sinaitic consonantary, which in turn led to the earliest Arabic consonantaries. Millard (1995: 193-94) reports the use of Nabataean numerals on the pre-Islamic Arabic inscriptions from En-Namara (dated 328 AD), and possibly on the 6th century AD Zabad and Harran inscriptions. Such late occurrences become increasingly rare, however, as Greek alphabetic numerals and systems based thereupon came to predominate throughout the Middle East, and by the time of the introduction of Islam, no trace of the Nabataean or any other Levantine system remained.

Kharoshthi

The Kharoshthi script was used in what is today eastern Afghanistan, northern Pakistan, and the Jammu and Kashmir state in northern India - altogether, the area known as the Hindu Kush - from around 325 BC to 300 AD in Aśokan, Saka, Parthian, and Kusana inscriptions. It is generally assumed by most Western scholars to be of Aramaic origin, given its proximity to the Seleucid Persians (for whom Aramaic was a *lingua franca*), the similarity in form and value of many of the signs in the two scripts, and the fact that both were written from right to left. Thus, although Kharoshthi inscriptions do not appear until somewhat later, it is generally accepted that contact with users of Semitic scripts was the immediate context in which the script was transmitted from west to east. The assumption of a Semitic origin is debated by some Indian scholars, many of whom see an independent origin for Kharoshthi, but this issue is not nearly so contentious as that of the origins of the Brāhmī script (the ancestor of all modern Hindu scripts). While Kharoshthi flourished for some time, it never achieved the popularity of Brāhmī, and it ceased to be used around 300 AD.

During the earliest periods of its use (before about 100 BC), Kharoshthi inscriptions containing numerals are quite rare, being found in only a few royal inscriptions of the Mauryan King Aśoka, who reigned from about 273 to 232 BC. Only the numbers 1, 2, 4, and 5 are represented, and they are always formed using simple unit-strokes. In the later Saka, Parthian, and Kusana inscriptions (dating from about 100 BC onwards), a more complex system was used, and much larger numbers were represented. This system possessed unique signs for the numbers 1, 4, 10, 20, and 100, as shown in Table 3.7.

Table 3.7: Kharoshthi numerals

1	4	10	20	100	
𐩀	𐩃	𐩄	𐩅	𐩆	𐩆

In common with all the Levantine systems, Kharoshthi is purely cumulative-additive up to 100 and multiplicative-additive thereafter. Unlike other Levantine systems used at the time, Kharoshthi has no special sign for 5; numbers from 4 through 9 were always expressed through combinations of units and 4-signs. Mangalam (1990: 48) reports the use of a special sign for 1000 (𐩇) in the Kharoshthi inscriptions found in Chinese Turkestan, but I have been unable to confirm this. As in the script as a whole (and in other Semitic scripts), the direction of writing is always from right to left. Thus, 697 would be expressed as 𐩀𐩀𐩀𐩃𐩄𐩅𐩅𐩅𐩅𐩆𐩀𐩀𐩃.

The earliest Kharoshthi numerical notation system, being formed solely with vertical strokes, need not have been of Aramaic origin, though the geographical proximity of its users to Seleucid Persia, coupled with the obvious relation of the Kharoshthi alphasyllabary to the Aramaic consonantary, suggests that it might have been. In its fully developed form, however, it is clearly part of the Levantine family. Kharoshthi shares with the other systems the right-to-left direction of writing (as opposed to Brāhmī), the use of vertical strokes for units, similar forms for the numeral-signs for 10 and 20, and the use of the multiplicative principle for 100. The use of 𐩃 for 4 is common to both Kharoshthi and Nabataean, which seems unlikely to be coincidental, since they share a common sign for 20 as well. The question of the origin of the symbol 𐩃 symbol is a thorny one, since both systems began to be used around 100 BC. As I mentioned already, the Hebrew hieratic sign for 4 was + or X, suggesting transmission from west to east. However, Datta and Singh (1962: 23) argue that the sign may have developed by rotating the Brāhmī sign for 4 (+) by 45°, which may then have been transmitted westward to the Nabataeans. Buhler (1896: 73), however, contends that the Nabataean and Kharoshthi signs were invented independently of one another. Whatever the case, the remainder of the Kharoshthi system is clearly derived from earlier Levantine systems, specifically Aramaic, so that the Levantine origin of most of the Kharoshthi system's structure is evident.

Kharoshthi numerical notation was used on various inscriptions (not to my knowledge on ostraca or in any medium other than stone), but was always in competition with its rival, Brāhmī, which was the script of choice of the Mauryan kings of the Indian heartland. The use of Kharoshthi was tied to the political independence of the Bactrians and Scythians, who looked to Greek and Persian traditions rather than Indian ones. By the late 3rd century AD, the polities of the Hindu Kush were seriously weakened, and the advent of the Gupta Empire in the 4th century AD signalled the end of Kharoshthi and the predominance of scripts (and numerical notation) descended from Brāhmī throughout the Indian subcontinent.

Hatran

In the early years of the Christian era, a variant Aramaic script was used in the region around the city of Hatra (modern Al-Hadr, in northern Iraq), an outpost of the Parthian Empire and later the capital of the small autonomous state of Araba. The Hatran script, for which inscriptions have been found dating from about 50 BC to 275 AD, possessed a distinct numerical notation system with signs for 1, 5, 20, and 100, as shown in Table 3.8.

Table 3.8: Hatran numerals

1	5	10	20		100
𐣻	𐣼	𐣽	𐣾	𐣾	𐣿

As with all Levantine systems, the Hatran numerical notation is decimal, cumulative-additive for numbers less than 100, multiplicative-additive above 100, and written from right to left. Thus, 697 would be written as 𐣼𐣻𐣿𐣾𐣾𐣾𐣾𐣽𐣼𐣻𐣻. The

Hatran system is clearly part of the Levantine family, though its precise relation to the other systems is unclear. It is obviously descended in some way from the Aramaic system that dominated the region around Hatra in the centuries prior to the development of the Hatran script, given the similarity of signs for 1, 10, and 20 to earlier Aramaic forms. The sign for 5 is identical to that of the Estrangelo script used around Edessa at that time, but is not obviously related to any other script. Finally, the 100-sign is of entirely mysterious origin, though a case could be made that it is related to the Phoenician .

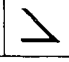
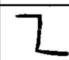
Like all Levantine systems, Hatran numerical notation was used exclusively for non-mathematical purposes, on ostraca and inscriptions and in certain economic documents. Unlike the Palmyrene and Nabataean states, both of which were subordinate to Roman political and economic domination for most of their history, Hatra remained independent from both Roman and Parthian control until 272 AD, when the Sasanian king Shapur I conquered the region. After this time, Hatran inscriptions are more rarely encountered. However, numbers appear on various Sasanian inscriptions dating from the 3rd to 6th centuries AD that seem to be derived from the Hatran system, although without a sign for 5 (Frye 1973). The study of Sasanian and later Pahlavi numerals remains very poorly understood, although it appears that the older cumulative Levantine signs became reduced over time into a quasi-ciphered form.

Syriac (Estrangelo)

A consonantal script was used at the ancient city of Edessa (modern Urfa, in southeast Turkey) in the early years of the Christian era. Based on an Aramaic model, this script, which came to be known as *Estrangelo* or *Estrangelâ*, was used exclusively by Syrian Christians until around 500 AD, at which time it diverged into western (Serto) and eastern (Nestorian) variants. A large number of Estrangelo inscriptions dating from 50 to 500 AD have been found in northern Syria and southern Turkey.

The Estrangelo numerical notation system possessed unique signs for 1, 2, 5, 10, 20, 100, and 500, as shown in Table 3.9.

Table 3.9: Estrangelo numerals

1	2	5	10	20	100	500
|				o		÷

Numeral-phrases, like the script itself, are always written from right to left. The system is decimal, cumulative-additive for numbers less than 100, and multiplicative-additive for higher values. The Estrangelo numerical notation is thus clearly part of the Levantine sub-group, given the direction of writing, the 20-sign and the use of multiplication. However, it has some curious features. The sign for 2 should probably be seen as a ligatured form of two unit-strokes, and may have begun as a paleographic convenience that only later became a structural feature of the system (Duval 1881: 14-15). The sign for 5 is identical to that of the Hatran system, the sign for 20 to one variant form used in Phoenician, and that for 10 identical to those used throughout the entire family. To complicate matters further, Duval (1881: 14) argues that the 100-sign () is a slightly modified form of the 10-sign, resembling in this respect the Palmyrene numerical notation. At present, the best that can be said of its origins is that Estrangelo numerical notation is a variant of the Aramaic system.

A unique feature of the Estrangelo numerical notation system is that it is the only Levantine system to possess a symbol for 500. The sign for 500 is rarely encountered, as numbers of this magnitude are infrequent in Syriac writings. Duval (1881: 14) insists that it ought to be understood in many numeral-phrases where it is not written, but I remain unconvinced, and interpret such phrases according to the values of the expressed signs. I suspect that the notion of a sign for 500, if not its form, may have been borrowed from the Romans, under whom the Syrian Christians remained subjugated throughout this period.

Estrangelo numerical notation was used on a wide variety of inscriptions, economic documents, and certain other manuscripts, though for liturgical and literary purposes numbers were written out in full. Although the Edessan Christians were politically subordinate to Rome throughout most of the history of their script and numerical notation, both survived several centuries of domination. Eventually, however, the numerals began to be displaced by the ciphered-additive Syriac alphabetic system (ch. 5), which assinged numerical values to the 22 letters of the Syriac consonantary. By around 500 AD, two separate Syriac scripts – a western (Monophysite) and Eastern (Nestorian) had begun to develop, but by this time the Estrangelo system had been entirely superseded.

Summary

The Levantine family is thus descended from the Aramaic and Phoenician systems developed around 750 BC. It seems to have been based on the dual model of the Egyptian hieroglyphic system and the Assyro-Babylonian common system. Over the second half of the first millennium BC, the Aramaic system and its descendants spread throughout Assyria, Persia, Egypt, Asia Minor, and even into India. By 300 AD, however, it was used only in pockets of the Middle East and in northwest India. While these systems were used for a full millennium, they ceased to be used once the polities in which they predominated (most significantly Achaemenid and Seleucid Persia) declined in importance. No numerical notation system descended from this sub-group survived past the rise of Islam.

The central features of Levantine numerical notation systems are as follows: a) an overall decimal system; b) a special sign for 20 (sometimes a combination of two 10-

signs); c) the use of vertical strokes for units and horizontal strokes (usually with some degree of curvature) for tens; d) a cumulative-additive structure for numbers less than 100; and e) use of multiplicative-additive notation for expressing multiples of 100 (and also 1000 and 10000 where appropriate). Signs for 4 are found in Nabataean and Kharoshthi, with a sign for 5 in late Aramaic, Palmyrene, Nabataean, Hatran, and Estrangelo.

While these systems were used extensively for administrative and mercantile purposes, as well as on inscriptions, there is no evidence that any Levantine numerical notation system was ever used as a computational aid either for arithmetic or for higher mathematics. It is furthermore essential to reiterate that in the various scripts in question, numerals were usually written out in full in most religious and literary contexts and a sizeable minority of economic documents and inscriptions. As such, numerical notation occupied a relatively minor position in these ancient Levantine societies.

Chapter 4: Italic Systems

Other than our own Western numerals, the Roman numeral system is certainly the best-known numerical notation system in most Western countries. It is less well-known, however, that the Roman numerals themselves are part of a larger family of numerical notation systems which originated on the Italian peninsula and that flourished between 500 BC and 500 AD, and which includes many systems used throughout the Mediterranean region, Western Europe, and North Africa. The cultural and political imperialism of the Greeks and Romans during this period spread many institutions, including this particular type of numerical notation. The Italic family of numerical notation systems includes the Etruscan, Roman (the classical system and its multiplicative, positional, Arabico-Hispanic, and calendar variants), Greek acrophonic (and its non-acrophonic variants), South Arabian (Minaeo-Sabaean), Lycian, and Berber systems. It is so named due to its origin on the Italian peninsula, and clearly not because of any linguistic or cultural affiliation shared by its users. These systems are summarized in Table 4.1, listing only the most common versions of the numeral-signs.

Table 4.1: Italic numerical notation systems

System	1	5	10	50	100	500	1000	5000	10000	50000	100000
Berber	I	>	O	⊏	b	X	⧖				
Etruscan	I	Λ	X	𐌣	Ж	ↀ	⊕	φ	Φ		
Greek acrophonic	Ι	Γ	Δ	𐅄	Η	𐅅	Χ	𐅆	Μ	𐅇	
Greek – Argos/Nemea	●		O	𐅄	日		X				
Greek – Epidaurus	●		—		日		X				
Greek – Olynthus	I		X		8		ᛘ				
Greek - archaic non-acrophonic	I	Γ	↑	⸢	↥						

Lycian	I	∠	O	Γ	IH						
Roman	I	V	X	L	C	D	ↀ	ↁ	ↂ	ↇ	ↈ
Roman multiplicative	I	V	X	L	C	D	$\bar{I}$	$\bar{V}$	$\bar{X}$	$\bar{L}$	$\bar{C}$
Arabico-Hispanic	ɕ	ح	ع	ـس	لـس						
Calendar numerals	⊦	ľ	T								
South Arabian	I	Ψ	o	ᚦ	ß		ň				

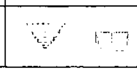
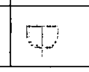
Despite the emphasis placed on Roman and Greek history by modern Western scholars, the history of this family remains relatively poorly understood. In particular, while the Greek acrophonic and Roman systems have received considerable attention, many systems used by politically peripheral peoples have been ignored in comparative studies. Part of this problem is due to a general lack of scholarly research, but these systems are also sometimes considered uninteresting because most of them are structurally similar to the well-known Roman numerals. Yet when the reconstruction of the systems' culture history is of great importance, these marginal systems are of great theoretical interest.

Etruscan

The Etruscans were a non-Indo-European people whose civilization had its centre in north-central Italy, in the region of modern Tuscany (whose name is taken from the Latin *Tusci*, meaning Etruscan). While the origins and civilization of the Etruscans are poorly understood, and large parts of their language remain undeciphered, there can be no doubt that the Etruscan civilization was the most potent political force on the Italian peninsula between around 800 and 300 BC. Furthermore, it had a very significant influence on Roman traditions throughout the Republic and even later. The Etruscan script, developed in the early 7th century BC, is alphabetic, representing both vowel and consonant sounds, and usually runs from right to left. The script was probably derived from alphabets of western Asia Minor that are closely related to that of archaic Greek.

The lexical numerals used by the Etruscans are a subject of some scholarly debate, but it is widely accepted that they were base-10 with a special term for 20, *zathrum* (but not for 40, 60, 80...) and in which subtractive structures were used to form the words for 17-19 (Lejeune 1981, Bonfante 1990: 22). However, these irregularities are not reproduced in the Etruscan numerical notation system, which is shown in Table 4.2.

Table 4.2: Etruscan numerals

1	5	10	50	100	500	1000	5000	10000
I	Λ	X	↑ ⋀	⋇ ƆIC C		⊕ ⨀		⏀ ⦶

This system is cumulative-additive with a mixed base of 5 and 10, very much like the Roman numerals, but is most often written from right to left, with the highest values at the right side of the numeral-phrase. Each exponent of the primary base (10) may be repeated up to four times, but the half-decade values may occur once only in any numeral-phrase. Thus, the number 1378 would be written as IIIΛXX⋀⋇⋇⋇⊕.

The numeral-signs for 1, 5, 10, and 50 are very well known and their forms are quite regular throughout the system's history. The sign ⋇ for 100, though less well known, is found in many inscriptions from a relatively early date; the use of C is seen by Keyser (1988: 542) as a development occurring between 250 and 200 BC. It may be that C = 100 arose first in Latin inscriptions and found its way into Etruscan inscriptions only when the Etruscan system was already declining. The signs for 1000 and 10000 are only encountered very rarely, in particular on the famed "abacus-gem" or "Etruscan cameo", which I will discuss below (Keyser 1988: 543, Fig. 8). The first form for each number in the chart is that described by Keyser (1988), while the second forms are taken from Bonfante (1990: 22). The signs for 500 and 5000 are entirely unattested, but are hypothetical reconstructions. Because the forms of the numerals 5 and 50 are the bottom halves of the signs for 10 and 100, we ought to expect this principle to be followed throughout the entire system: ⊕ = 1000 to ϖ = 500; ɸ = 10000 to Ψ = 5000 (Buonamici 1932: 244; Keyser 1988: 544-5). This might seem unlikely, but this theory is given some support in that the earliest Roman sign for 500 is Ɵ (later to become D, of course). Furthermore, there is some epigraphic evidence in that two previously unidentified Etruscan signs ϖ have been assigned such a value (Keyser 1988: 545). The highly speculative assertions of De Feis (1898, reproduced in Buonamici 1932) concerning Etruscan signs for 50,000 and 100,000 ought not to be given any serious consideration. A special sign, >, is used primarily on coins to indicate ½ (Bonfante 1990: 48).

While the Etruscan script is attested from around 700 BC or shortly thereafter, there is no evidence to my knowledge of the use of the numerals until the late 6th century BC. Yet our chronological knowledge is severely limited, making culture-historical reconstructions of the origin of the system difficult at best. There are three general theories for the origin of the Etruscan numerals: independent invention based on tally-marks, diffusion from the Mycenean numerals, and diffusion from the Greek colonies of southern Italy.

The first theory holds that their invention was entirely independent from other base-10, cumulative-additive systems of the Mediterranean that existed in the early first millennium BC. This theory would be greatly supported if it could be established that the Etruscans were indigenous to the Italian peninsula and were not descended, as others claim, from a non-Indo-European people of Asia Minor, although this is not a necessary condition for the independence of the system. It is strengthened by the fact that no earlier system used a mixed base of 5 and 10 (although see below for discussions of the Greek acrophonic system). The best evidence for this theory comes from the form of the numeral-signs. Most of the signs can be derived from successive crossings and circlings of tally-marks for 1, 5, and 10. The most common sign for 100 is simply 10 with a vertical line through it, while 50 is made by drawing a straight line from the apex of the "upsidedown V" 5-sign (Keyser 1988: 533). The concept of tallying numbers, which is basically an ordinal technique where numbers are marked sequentially, then crossed off as appropriate, could thus have led to a cardinal numerical notation system. The main problem with this theory is that because tally-sticks are normally wooden, we have no surviving evidence that the Etruscans ever used tallies in such a manner.

The second theory holds that similarities between the Mycenaean Linear B numerals and those of the Etruscans are the result of cultural contact between the two groups at some point in their history. This is not as far-fetched as it might sound, given that we do not know very much about the geographic origins of the Etruscans. The theory that some aspects of Etruscan culture may be indebted to a Mycenaean Greek influence has been most powerfully argued by Peruzzi (1980) who, however, does not discuss the numerical evidence. It is possible (though unproven) that the Etruscans migrated from western Asia Minor during the Aegean "Dark Age" between approximately 1100 and 800 BC. Mycenaean settlements have been found in southern Italy and Sicily, though these are really too early to have had much cultural contact with the Etruscans. Keyser (1988: 542-3) notes that the "Minoan-Mycenaean" numerals (he ignores the differences between the Linear A and Linear B systems), like Etruscan, use strokes alone to represent the lower exponents of the base and strokes in conjunction with circles for higher exponents. Unfortunately, this theory suffers from the defect that 100 is Ж in Etruscan but ◎ in Mycenaean. The similarity between Etruscan ⊕ and Mycenaean ⊖ for 10,000 is notable, but it is the only numeral-sign to be relatively close in form and value in both systems, other than their common use of a vertical stroke for 1, which is historically nearly meaningless. Haarman's (1996) theory advocating Aegean ancestry for the Etruscans based on their numerals ought to be discounted on the basis that he compares sign forms alone, while ignoring the fact that the meanings attached to them are entirely different. Although Ж is found both in Etruscan and Aegean representational systems, this tells us nothing of value, because it is a numeral-sign for

100 in Etruscan but represents the syllable [a] in the Cypriote syllabary. At any rate, the evidence from sign-forms alone is not enough to decide the case. Evidence from other aspects of culture still needs to be compiled before deciding this question.

A third theory, which, surprisingly, has not received great attention, is that the Etruscan numerals may be related to the Greek acrophonic system described below, though whether this relation is as ancestor or descendant awaits the establishment of better chronological sequences. Admittedly, the numeral-signs of the two systems are quite different, except for the use of **I** for the units. Otherwise, though, this theory has considerable support. The two systems both use a sub-base of 5, which is not found in any other system at that time or earlier. The ancestral role of the Greek scripts with respect to Etruscan is now very widely accepted, and many other aspects of Etruscan culture owed much to contact with Greek traders. There is further direct evidence of a connection between the two. Alan Johnston points out that many early Greek numerals are found in the late 6th century BC in south Italy in the context of contact with the Etruscans (Johnston 1975: 362-364; Johnston 1979: 31). Moreover, tantalizingly, the use of **X** for 10 is occasionally found in these early Greek systems, before the use of the acrophonic principle ended its use in this manner. Unfortunately, this evidence of cultural contact and similarity of structure shows only that a relationship probably exists, but cannot establish whether the Etruscan or the acrophonic system has priority. At present, our knowledge of the chronology of the two systems indicates only that both were probably developed in the mid- to late 6th century BC on the Italian peninsula.

My general evaluation of the merits of these three theories is that the Etruscan system probably arose relatively independently of other systems, possibly with some continuity or influence from Mycenaean (Linear B) numerals. A theory arguing for total independence would have to cope with the fact that base-10, cumulative-additive systems abounded in the Mediterranean throughout the period between 1100 and 650 BC. The Egyptian hieroglyphic system used in the Nile valley, the Aramaic and

Phoenician systems of the Levant, the Hittite hieroglyphic numerals, the Cypriote numerals, and possibly remnants of the Mycenean system were all used in this period and thus may have been known to the Etruscans. The theory of Mycenaean origin is interesting. Nevertheless, while this theory explains the similarities in the signs for 1000 and 10000 in the two systems, it remains essentially unproven until more can be determined regarding the geographical origins and cultural affiliations of the Etruscans. Yet the invention of a mixed base of 5 and 10 (thus reducing the number of times any one sign need be repeated from nine to four) is an important creative event, one that at present seems to rest with the Etruscans. Also, the use of halved signs for the sub-base of 5 is an ingenious means of deriving sign-values, and suggests that whatever system(s) the Etruscans knew, their numeral-signs are of their own invention. I am not convinced of the temporal priority of the Greek acrophonic system over the Etruscan. Instead, the evidence seems to suggest the transmission of the numerals from the Etruscans to the Greeks to be more likely, and I thus reject the alternative possibility unless further evidence is forthcoming.

The Etruscan system is clearly the direct ancestor of the Roman numerals. This should come as no surprise, given that Etruria was politically dominant over Rome throughout its early history and remained a potent force in Roman politics well into the Republican period, long after its supremacy on the Italian peninsula had ended. I will discuss the indebtedness of the Roman numerals to the Etruscan system in detail below. In addition, a number of other Indo-European languages of the Italian peninsula, including Oscan, Umbrian, and Faliscan, adopted scripts and numerical notation systems based on an Etruscan model, although these numerals are only encountered rarely and are essentially identical to the Etruscan numerals. Finally, if the Etruscan system is temporally prior to the Greek acrophonic system in southern Italy, it is extremely likely that the Greek colonists borrowed the system from the Etruscans, adopting the structure of the system while using the acrophonic principle to adapt the numeral-signs to their own language.

Etruscan numerals were used in a wide variety of contexts. A great deal of evidence from funerary inscriptions has survived in which numerical notation is primarily used to indicate the age of the deceased; other inscriptions on stone make use of numerical notation, but more rarely. From the 5th century BC onwards, Etruscan coins are found using the numeral-signs for ½, 1, 5, 10, 50, and 100 in various combinations. As well, the graffiti inscribed on potsherds contain many instances of Etruscan numerals. These graffiti were used on containers for recording the quantity of goods contained therein or their values. A lead tablet whose purpose has not been reliably established is notable in that it contains the numeral-signs for 1000 and 10000 (Keyser 1988: 544, Fig. 9). Finally, but perhaps most interestingly, the "Etruscan cameo" or "abacus-gem" is a small gem (1.5 cm high) which depicts a seated man working at what can only be an abacus – a large board upon which rows of Etruscan numerals have been inscribed, including the elusive signs for 1000 and 10000, but not 500 or 5000 (see Keyser 1988: 545). This not only is the earliest known use of any numerals on jewellery, but also demonstrates the association of the numerals with pebble-board computation. Of course, it is also quite plausible to postulate the use of the Etruscan numerals on wooden tallies and similar perishable materials, despite the lack of evidence for such a function. Unsurprisingly, there is no evidence for the use of the Etruscan system for performing arithmetical calculations as we would (on paper or slate). Computations would have been done in the head, with the fingers, or on a counting-board. Despite the existence of the lead tablet and the Etruscan cameo, large numbers and long numeral-phrases are very rarely encountered; even the sign for 100 is relatively uncommon.

The demise of the Etruscan numerical notation system was a direct consequence of the rising fortunes of the Roman republic. The lack of Etruscan political unity in the 3rd century BC, coupled with the undeniable political advantages of association with

Rome, led to the slow but steady assimilation of the cities of Etruria into the Roman political and cultural milieu. While the Etruscans remained a culturally distinct people at least until the beginning of the Roman Empire, by 100 BC, they were entirely within the Roman political sphere. This inevitable trend was accompanied by the slow replacement of the Etruscan language, script, and numerical notation with those of the Romans. Given that the Etruscan system was directly ancestral to the Roman numerals and was very similar to it, there would have been little difficulty in making the change to the new system. While there are some 2nd century BC examples of Etruscan numerical notation, these are among the latest examples known. Or are they?

Tuscan Tallies: A Modern Survival?

A curious theory concerning the survival of Etruscan numerals until the present day was first outlined by A.P. Ninni in the 19th century (Ninni 1888-89). While studying the tally-marks used extensively at the time by fishers along the coast of the Adriatic Sea in Tuscany, the homeland of the Etruscans, Ninni discovered a numerical notation system known by the people as *cifre chioggotte*. This potential vestige is cumulative-additive, with a mixed base of 5 and 10, like both the Roman and Etruscan systems. Its numeral-signs are shown in Table 4.3.

Table 4.3: Tuscan tally numerals (Ninni 1888-89: 680)

1	5	10	50	100	500	1000
I	Λ V C	X I ⌐	ᛏ A V	Ж ⊙ ⊕	ЖΛ ᛏ	M ✳ [illegible]

Ninni noted that the signs of the *cifre chioggotte* more closely resemble the Etruscan numerals than the Roman numerals, and was well aware of the geographical and cultural connection implied by their presence in Tuscany. He further noted that in both the Etruscan and Tuscan systems, several of the sub-base numeral-signs were halved versions of the signs for exponents of ten (Ninni 1888-89: 680-1). Could this system in fact be a survival, over 2000 years, of an Etruscan tradition among modern Tuscan peasants? Ifrah (1985:150) believes the level of continuity to be even greater, seeing this and other European tally systems as an enormous cultural substratum to writing lasting from the mists of prehistory to the present day, of which the Etruscan and later Roman numerals were merely offshoots.

At present, there is simply not enough surviving data to speculate on the possibility of such long-term cultural survivals, particularly in a region, such as central Italy, that most certainly has not been a cultural backwater for over two millennia. I admit that, if Ninni's data are right, certain signs of the *cifre chioggotte* (e.g. the first signs in the above table for 50 and 100) are identical to the Etruscan numeral-signs for those numbers, but are quite dissimilar to the intervening Roman numerals, which also probably played a role in their development. Since no new evidence has become known on this subject for over a century, perhaps we have lost our opportunity to learn more about this system.

Roman

The history of the Romans is so well known as to make its discussion here almost irrelevant. From its roots as a tiny Italic city-state under Etruscan domination, to its control of the entire Mediterranean region and beyond, to its continuing influence on matters of language and law even today, Rome's influence on Western history is nearly inestimable. Nevertheless, despite the importance and continuing use of Roman numerals to the present day, we know far too little about their origin. In fact, until Keyser's (1988) recent study, the origin of the Roman numerals was very poorly understood indeed, with misconceived reconstructions promulgated as fact by generations of classicists. Similarly, it is not normally recognized that there are several different structural variants of the Roman numerals, and that the sign forms and structure of the Roman numerals used today are hardly older than the Western numerical notation system.

The Roman alphabet was developed on an Etruscan model around 600 BC at a time when much of central Italy was under Etruscan political domination; it was written from left to right, as it is today. The Latin lexical numerals are decimal in structure and use the subtractive principle for 18 and 19 (*duodeviginti, undeviginti*). As with the Etruscan numerals, the classical Roman numerals do not reproduce this irregularity. Like its Etruscan precursor, the Roman system has a base of 10, with an auxiliary base of 5, and is largely cumulative-additive in structure (but see below); unlike it, the Roman numerals are written from left to right and are sometimes used subtractively.

The variety of numeral-signs used throughout over two millennia of the history of this system is astonishing, particularly in comparison to the highly static quality of the equally long-lived Babylonian cuneiform and Egyptian hieroglyphic numerals. Table 4.4 presents the vast majority of numeral-signs used during the republican period.

Table 4.4: Roman numerals (republican period)

1	5	10	50	100	500
I	V Λ	X	Ѱ ↓ ⊥ ⊥ L	C Ɔ	D Ð
1000[1]	5000	10000	50000	100000	
ↀ ∞ Φ ⋈ ᔕ ⅄	ↁ ħ	⅄ ↂ ⩛	ↇ	ↈ ↈ	

Of all the numeral-signs used in republican inscriptions, only the signs for 1 and 10 remain unchanged throughout the entire history of the system. The "inverted-V" sign for 5, Λ, is found only in early contexts, and is evidence of the system's indebtedness to its Etruscan ancestor, but V is used exclusively thereafter. The signs for 50 in Table 4.4 are roughly in chronological order; Ifrah (1985: 149) describes their evolution from

[1] See Ifrah 1985: 132 for many other forms for 1000 used in republican and early imperial Rome.

inverted forms of the Etruscan numeral-sign for 50 to their final form as the letter L, which was firmly established by around the 2nd century AD. The Ↄ form for 100 is extremely rare; Ifrah (1998: 188) lists only a single inscription where it is found. Keyser indicates that the first Roman C=100 whose date is secure is from 186 BC, but he regards a 3rd century BC origin for the symbol as likely, even though there is no definitely datable example of any sign for 100 at this early date (Keyser 1988: 542). 500 is expressed using Ɵ in all early contexts, with assimilation to the alphabetic D occurring around the transition to Empire. The familiar M=1000 is never used until the Middle Ages, except, as discussed by Gordon (1983: 45), in various places where M is not used in numeral-phrases but simply as an abbreviation for *mille*. The signs for 5000, 10000, 50000, and 100000 are rarely encountered, though they are all attested as early as the 3rd century BC. The most likely principle governing their formation is that adding arcs on either side of the most common sign for 1000, ↀ, indicates successive exponents of 10, while the right half of the appropriate base-10 sign represents the quinary component: ↀ(1000)→D(500); ↂ(10000)→ↁ(5000); ↈ(100000)→ↇ (50000).

In addition to these symbols, two other symbols are found more rarely and do not fit neatly into the system described above. The first is a special sign for 6, ϛ, which is nothing more than a cursively written and ligatured *vi* or *ui*. It occurs on many late classical inscriptions from the 2nd and 6th centuries AD (Gordon 1983:46) and, on 6th century Byzantine imperial coins, ϛ was as common as VI (Wroth 1966: cx). Its use likely died out in the eighth century (Bischoff 1990: 176). The second outlying sign, Ꝗ, is used to represent 500,000 in a very limited number of examples (Mommsen 1965[1909]: 788-791; Gordon 1983:45). The sign is probably derived from alphabetic Q and is thus an abbreviation of *quingenta milia*. Its use was limited to the later Republic, and it was certainly not familiar to Pliny the Elder a century later, who, in his Natural History, wrote "*Non erat apud antiquos numerus ultra centum milia*" or "Among the ancients there was no numeral larger than 100,000" (Natural History 33.47.133).

These numerals had a purely cumulative-additive structure in most inscriptions, thus, 19494 might be expressed as ↂↁↀↀↀↀCCCCↆXXXXIIII. Around the late republican period, however, two changes began to occur. Firstly, the basic structure of the numerals was supplemented by the use the subtractive principle for multiples of 4 or 9 of the exponent. In these expressions, placing a lower-valued numeral-sign to the left of a higher one indicates subtraction of the former from the latter. This is done to reduce the overall length of numeral-phrases - where 4 or 5 numeral-signs would have been needed, only 2 are required. Thus, 19494 = ↂↀↂCDXCIV, a reduction from 19 to 9 numeral-signs. The use of addition alone certainly precedes the use of subtraction; it is used almost exclusively in the earliest inscriptions and is the more usual form even in later classical inscriptions (Sandys 1919: 55-56).[2] Cajori notes that its use is extremely rare before the Renaissance, except where a numeral is placed at the end of a line of an inscription (Cajori 1928: 31). Presumably, this allowed the engraver to include the entire numeral-phrase on a single line without needing to crowd many numeral-signs into a limited space. Whether this feature is related to the use of subtraction for certain lexical numerals in Latin remains unclear. However, it is clear that Guitel's denigration of the subtractive principle on the basis that it led the Romans away from the pure and easily understandable additive principle merely to improve the conciseness of the system is quite unwarranted (Guitel 1975: 202-3). There is no evidence that the Romans or any later users of the system found the subtractive principle unwieldy or difficult to understand, and it ought to be regarded as a very economical way of structuring numeral-signs even though it detracts from the "purity of principle" of the additive Roman numerals.

The second change, perhaps more important, is the introduction of the multiplicative principle when expressing very high numerals. Even as early as the 3rd

---

[2] It is curious that in the tradition of modern Roman numeral hour-numbers on clocks, 4 is normally denoted additively (IIII), while 9 is denoted subtractively (IX).

century BC, the Roman republic had become a large centralized state, and the need to express large numbers was acute. The highest number expressible with a single sign at the time was 100,000. At times, this led to extremely cumbersome numeral-phrases, such as the inscription on the famed *Columna rostrata* erected in Rome in 260 BC and re-cut in the first century AD, which celebrated a naval victory over Carthage in which over two million *aes* worth of loot were plundered. The column is inscribed with at least 22 ↈ signs for 100,000, and possibly as many as 32, as the inscription is somewhat fragmentary (Menninger 1969: 43-44). One is struck, in looking at this inscription, at the sheer enormity of the numeral-phrase, and thus, by the impressive amount of booty obtained.

By the end of the Republic, the principle of multiplication began to be used to express multiplication by 1000; a horizontal bar (*vinculum*) was placed above a numeral-phrase or some portion thereof to indicate that the number under the bar should be multiplied by 1000 to get its true value. However, for many smaller numbers, multiplication did not improve conciseness; to express 191063, one might write $\overline{\text{CLXXXXI}}\,\text{LXIII}$ instead of ↈↇↂↂↂↂↀLXIII, but twelve symbols are still necessary. The main advantage in this case is that one need no longer remember so many numeral-signs or invent new ones; the signs for 1, 5, 10, 50, and 100 are sufficient to express any number up to 500,000, whereas eleven different signs would be needed under the purely additive system. Gordon (1983: 47) indicates that this principle is first used in the *Lex de Gallia Cisalpina* written between 49 and 42 BC. Within a very short time (perhaps as little as 25 years), early Imperial Romans found it necessary to express even higher values; as a result, three vertical bars enclosing a numeral-phrase on the top and sides signified multiplication by 100,000. Thus, instead of the 32 signs for 100,000 found on the *Columna rostrata,* one would only need to write $\left|\overline{\text{XXXII}}\right|$. This principle was first used in the late 1st century BC, according to Gordon (1983: 47), but was employed rather sparsely until the 2nd century AD. By the early Empire, any number less than 500 million could be expressed with just the five lowest numeral-signs plus two types of bar to express multiplication. This revised system is still a decimal system with a sub-base of 5; however, instead of being purely cumulative-additive, it is a hybrid system using cumulative-additive structuring for numbers up to 1000 and multiplicative-additive thereafter. The entire system (up to 100,000,000) as used in the Imperial period is shown in Table 4.5.

Table 4.5: Roman numerals (Imperial period)

	1	5	10	50	100	500	1000
Regular signs	I	V	X	L	C	D	ↀ
	1000	5000	10,000	50,000	100,000	500,000	1,000,000
Multiplicative (1000)	$\overline{\mathrm{I}}$	$\overline{\mathrm{V}}$	$\overline{\mathrm{X}}$	$\overline{\mathrm{L}}$	$\overline{\mathrm{C}}$	$\overline{\text{Ð}}$	$\overline{\text{ↀ}}$
	100,000	500,000	1,000,000	5,000,000	10,000,000	50,000,000	100,000,000
Multiplicative (100,000)	$\overline{\vert\mathrm{I}\vert}$	$\overline{\vert\mathrm{V}\vert}$	$\overline{\vert\mathrm{X}\vert}$	$\overline{\vert\mathrm{L}\vert}$	$\overline{\vert\mathrm{C}\vert}$	$\overline{\vert\text{Ð}\vert}$	$\overline{\vert\text{ↀ}\vert}$

Gordon (1983: 47) claims that the largest number expressed using this hybrid cumulative and multiplicative system is 35,863,120: $\overline{|\mathrm{CCCLVIIII}|}$ $\overline{\mathrm{LXIII}}$ CXX, though Menninger presents a photograph of an inscription from 36 AD that apparently indicates 100 million as $\overline{|\text{ↀ}|}$ (Menninger 1969: 245). Most of the higher signs are attested only rarely. Curiously, Guitel, who is particularly interested in the hybrid multiplication used in the Greek alphabetic (ciphered-additive) system, does not see the merit in this Roman invention, but instead regards it as an evolutionary "dead end" (Guitel 1975: 215). Because the Romans could now express very high numbers with very few symbols, she argues, they no longer needed to develop a more efficient positional system. The ethnocentrism and teleology of this argument is immediately apparent, as it regards this development only with respect to its failure to lead to a "superior" system. The Romans themselves likely perceived it as a means of improving conciseness, while reducing the number of signs one needed to memorize. Although the use of the 1000 "bar" continued among some post-Roman scribes, the use of the 100,000 "box" did not outlast the Empire.

The origin of the Roman numerals is one of the most hotly debated topics in the study of numerical notation, among both classicists and historians of mathematics. Unfortunately, as Cajori pointed out almost seventy-five years ago, "the imagination of historians has been unusually active in this field" (Cajori 1928: 31). Fortunately, Keyser's panoptic essay on the origin of the Roman numerals, which examines a variety of theories developed, ranging from the sixth-century theories of the grammarian Priscian to the twentieth-century theories of modern classicists, has, I believe, firmly settled the issue (Keyser 1988). While readers interested in this survey should consult Keyser's paper, I will briefly discuss widely held misconceptions about the Roman numerals.

It should be clear already that any theory of alphabetic origin of the numeral-signs must be rejected. While the modern Roman numeral-signs are also letters, even a brief glance at the other numeral-signs indicates that this cannot be so. While the signs for 1, 5, and 10 are indeed letters, one must engage in a great deal of special pleading to explain why the signs I, V, and X would be accepted rather than others: U, Q, and D, for *unus, quinque,* and *decem,* for instance. While C is, enticingly, the first letter of *centum,* this is probably a fortuitous coincidence, since Keyser (1988: 542) argues that C as 100 is a reduction of the older Etruscan sign Ж, which was sometimes also written as ƆIC. The signs for 50 and 500 only became associated with the letters L and D relatively late in the development of the system, certainly not before the late Republic, and M was not used for 1000 until the Middle Ages. It is fortuitous that the older sign for 1000 could be easily transformed into an M. Still, while the alphabetic character of the later Roman numerals cannot be invoked as an explanation of their origin, the fact that the numeral-signs all eventually came to resemble letters is of some importance. The alphabetic nature of the signs certainly served as a mnemonic aid (particularly for C=*centum* and M=*mille*) because one would not need to learn an entirely new set of signs for the purpose. Yet this is a later development and does not help explain the origin of the Roman numerals.

A second set of theories relates to the use of the pictographic principle to explain the form of the Roman numerals from various positions of the human hands. In most theories of this sort, I=1 is taken from a single finger; V=5 represents an outstretched hand, while X=10 represents two hands placed together. The vast majority of such theories were offered (and rejected) between 1655 and 1725, but their mention by the prominent classicists Mommsen (1965 [1909]) and Sandys (1919) has led to their survival. While this explanation is imaginative, there is no evidence to support it.

Thirdly, and most importantly due to its continuing importance, is a theory developed by Theodor Mommsen in the mid-19th century, promulgated by him for decades thereafter, and now accepted in most handbooks and texts on the subject (Mommsen 1965[1909]; see Keyser 1988:538 for a list of texts in which the theory appears). Mommsen argued that the signs for 50 (ↆ), 100 (C), and 1000 (ↀ) were taken from letters of the Chalcidic Greek alphabet which may have been a model for the Latin script, but which were not needed to transliterate the Latin language: chi, theta, and phi, respectively, which in the Chalcidic alphabet do indeed resemble the numeral-signs. Unfortunately, this attractive theory has several flaws: the sign for 100 does not really resemble the Chalcidic theta; these letters are sometimes used in Etruscan and Roman inscriptions; and there are other Greek letters (e.g. zeta and sampi) not needed or used in Roman inscriptions. A further fault is that special pleading is required to derive the origin of Ð=500. For these reasons, it is probably best to view this theory with skepticism, if indeed it is not better to dismiss it entirely.

The Roman numerals (at least those less than 1000) almost certainly developed through direct diffusion from the Etruscans. In particular, the epigraphic evidence collected over the past century has rendered theories such as Mommsen's rather obsolete; indeed, it is surprising that such speculations have survived so long. The astonishing similarity between the Etruscan and archaic Roman numeral-signs, as shown in Table 4.6, would itself be enough to prove a relationship between the two systems, their identical structure and the coexistence of the two societies in space and time being superfluous. It is clear that the Etruscan numerals have temporal priority over the Roman numerals, which do not appear until well into the 5th century BC, and are not frequently encountered until the 3rd century BC. In fact, one could argue that the similarities between the two systems are sufficient to warrant their consideration as a single numerical notation system; they are identically structured and many of their numeral-signs are similar or identical. I choose to treat them separately, in part because the two systems are written in opposite directions and in part because the Roman system begins to use signs for much larger exponents at a relatively early date.

Table 4.6: Etruscan and Roman numerals

	1	5	10	50	100	500	1000
Etruscan	I	Λ	X	⩚ ↑	C Ж	⩀	⊕ ⦂⃝
Roman Republican	I	Λ V	X	⩛ ↓ ⊥	C	Ɖ	ↀ Φ

The Roman numerals were used in a broader range of contexts than any other cumulative-additive system. In its earliest forms, the Roman numerical notation system was used on coins, on pottery, and on inscriptions dealing with a wide variety of topics. Dates, monetary values, and measures were all frequently expressed in the Roman numerals. The Roman numerals could be employed to express both cardinal and ordinal values. Their use in administration and literature was widespread from the republican period onward. In texts, Roman numerals were used to enumerate page and line numbers as well as serving many functions in the body of the text. Accounts, inventories, and legal documents also occasionally provide us evidence of their use in commercial and institutional contexts.

While Roman numerals certainly were used in the contexts of arithmetic and calculation, there is minimal evidence that they were ever used *for* calculation. Glautier (1972) discusses the Roman account-records, which he characterises as primitive from an accounting standpoint because of the lack of positionality. A similar point is raised by Meuret (1997), in his discussion of the Lamasba tablet, an irrigation regulation from North Africa during the reign of Elagabalus (218-222 AD), which is essentially a multiplication table to enable quick calculation of water supplies, thus overcoming the computational deficiencies of the system. By far the most common computational function for which Roman numerals were used, however, was simply to mark the rows on the Roman abacus. While we do not have very many surviving Roman abaci, Taisbak (1965) has made a strong claim that the Romans did all their calculations with them. I am unconvinced by his argument that "the notation of Roman numerals originates from the abacus reckoning", which is contradicted by the derivation of the Roman numerals from the Etruscan system and ultimately from an older tallying system (Taisbak 1965: 158). By virtue of the fact that cumulative systems use one-to-one correspondence intraexponentially, just as one counter equals one multiple of an exponent on the abacus, the Roman techniques of numeration (Roman numerals) and computation (the abacus) complement one another. Moreover, this correlation is confirmed by the quinary (base-5) component of the abacus (there are rows not only for the exponents of 10 but also for their halvings). In so doing, no row on the abacus ever would have contained more than four counters, which would have facilitated reading and working with them. However, because the original Etruscan system probably emerged from a system of tallying, it is more likely that the structure of the abacus emerged out of the structure of the numerals rather than vice versa. To evaluate the efficiency of Roman numerals for computation without considering their interaction with the highly efficient abacus is thus entirely unjustified.

Despite the enormous influence of Roman civilization on Europe, North Africa, and the Middle East, and despite the extraordinary chronological duration of the Roman numerals (almost 2500 years, given that the system is still in use today in limited contexts), it produced relatively few descendants. While other systems in use over similar periods, such as the Brāhmī numerals and Greek alphabetic numerals, changed their form greatly as they spread across time and space, the Roman numerals of antiquity spread largely unmodified throughout Western Europe and other areas where the Roman alphabet was used. While the Indian and Greek systems spread throughout many different scripts, changing the forms of signs as they were transmitted, Roman numerals were infrequently adopted by users of other scripts.

Of the few descendant systems of the Roman numerals, I have already discussed the hybrid multiplicative-additive system used occasionally from the 1st century BC onward. In the medieval period, the system was essentially the same as its classical antecedent, though with slight differences in form and structure. In Arab-influenced Spain, certain variant Roman numeral systems were occasionally used starting in the 10th century AD. Meanwhile, in northern Europe, certain types of medieval calendars contain unusual Roman numerals. Finally, as the Roman numerals came increasingly under assault from the rival Western system, certain positional variants of the numerals were occasionally used, combining features of both the Roman and Western systems.

Computation with Roman Numerals

The computational efficiency of the Roman numerals is a subject of considerable antiquity in the history of mathematics. The consensus of these arguments, with which I am in general agreement, is that the Roman numerals are poorly suited to performing arithmetical calculations. Even where Roman numerals were used in bookkeeping and commerce, they were never used, as our numerals are, for actually performing computations. Rather, in all such cases, abaci were used to actually manipulate the numbers, with the Roman numerals being used only to express the result.

Yet the rejection of the efficiency of the Roman numerals for computation has been used to make two implications that I do not feel are particularly warranted. Firstly, it is unlikely that the use of the Roman or other cumulative-additive numerals has any simple or unilinear correlation with developmental stages of psychology, as Murray (1978), Hallpike (1979), and Dehaene (1997) have suggested, either as the cause or effect of a less abstract way of thinking about number. Even if many of the earliest and independently invented numerical notation systems are cumulative-additive, we cannot assert on that basis alone that the Roman number concept is less abstract than our own. Other evidence, such as the use of abaci and finger counting, refutes such a simple correlation between the cultural evolution of numerical notation and cognition. There may well be some correlation between these two areas, but it must be demonstrated, not assumed from the inefficiency of the Roman numerals for a task for which they were never intended.

A second argument that must be addressed is that the inefficiency of the Roman numerals prevented the Romans from developing other useful institutions or techniques. Glautier (1972) has proposed that the failure of the Romans to develop an efficient accounting system can be directly attributed to the lack of a suitable numerical notation system. Yet the Romans clearly had sufficient accounting techniques to administer their empire for several centuries, and while double-entry bookkeeping could only arise where there was a ciphered-positional numerical notation system, it obviously must have functional equivalents, or else no society lacking such numerals could administer a large political entity. The inferiority of Roman numerals is used by Guitel (1975) to explain why the Romans were poor mathematicians as compared to their Greek subjects, who had the ciphered-additive alphabetic numerals at their disposal. However, both Greek and Roman mathematicians relied considerably on prior Babylonian discoveries, and educated Romans were fully aware of the Greek alphabetic numerals. If, in fact, the Romans were poorer mathematicians than the Greeks, other evidence must be sought.

A small body of research within the sub-discipline of the history of mathematics holds that the Roman numerals are not inefficient and that there is no reason to regard them as less suited for computation than the Western numerals. There have been at least four attempts by modern scholars to show how the Roman numerals could have been used in written calculations without the aid of an abacus or similar technology (Anderson 1956; Krenkel 1969; Detlefsen *et al.* 1975; Kennedy 1981). These analyses, apparently derived independently of one another, differ in the exact technique used in performing calculations – for instance, whether or not numeral-phrases are lined up as in Western-style computation or how subtractive forms are treated. Their conclusions, though, are the same: that even if the Romans never used their numerals in such a fashion, the Roman numerals are in fact amenable to computational functions. While this is superficially true, I regard this argument as highly spurious for at least four separate reasons.

I. Numeral calculation versus mental calculation. Properly arranged, one could probably do arithmetic using Roman numerals or any other system. Perhaps, as is claimed, the Roman numerals are even easier to use for addition and subtraction than our own numerals (Smith and Ginsburg 1937: 18, Anderson 1956: 148). Yet when one performs arithmetical computation with *any* system, the process is not solely dependent on manipulating the numerals themselves, but relies largely on mental arithmetic. In part, the "efficiency" of the Roman system rests on the fact that when numerical notation is used for computation, the numbers represented must still be converted into mental concepts. When we do arithmetic, we do not use numerical notation alone; rather, lexical numerals and memorized tables of facts contribute to our computational ability. It is equally unclear whether the authors who have attempted to do arithmetic using Roman numerals in fact mentally translated the Roman numerals into mental representations in Western numerals; if so, the entire process is tainted.

II. Failure to compare systems. Many of the studies that criticize the Roman numerals' efficiency for computation assume that the system's inefficiency led to its replacement by the Western numerals. This occurred, not coincidentally, between the

13th and 16th centuries, at a time when both capitalistic commercial institutions and formal mathematics were growing rapidly. None of the abovementioned authors demonstrate that the Roman system is *equally* or *more* efficient than the Western or any other system. I am not sure how one would do such a comparison, since every child in the West is bombarded with Western numerals from infancy onward. It would be difficult for anyone today to become equally familiar with the Roman and the Western numerals.

III. Complex computational techniques. The techniques proposed by the abovementioned authors are in fact more complex than they are presented to be, and require considerable knowledge of mathematics and substantial mental calculation. For instance, the so-called "simple" calculations proposed by Detlefsen *et al.* involve a complex transformational "grammar" which could not possibly have been undertaken with the knowledge possessed by Roman or medieval scholars. One cannot simply take the Roman numerals as a set of signs independently of knowledge about how to combine or use them. These analyses are interesting mathematical exercises, but their complexity makes it unlikely that they were used as described by the Romans.

IV. Ahistorical nature of argument. All the abovementioned studies recognize that there is no evidence that the Roman numerals were ever used for arithmetic, but ignore this in favour of hypothetical calculation techniques. Anderson deals with the historical lacuna by suggesting that "any reader, once he discovers how simple the operations are, will be inclined to imagine that some Roman engineers and surveyors, in building their great projects, did occasionally do their computations very much in the way described below, even though they left no records of their work" (Anderson 1956: 145). Detlefsen *et al.* go so far as to blame the Romans for not recognizing the potential of their system for doing arithmetic (1975: 147). All incorrectly imply that arithmetical calculation is the primary or most obvious function of numerical notation. While the possibility cannot be ruled out that the Roman numerals were used in this way, the argument *ex silentio* is highly implausible, particularly given that we have persuasive evidence that the Greeks, Romans, and later medieval scholars used the abacus to do computations (Lang 1957; Taisbak 1965; Murray 1978). Such arguments are not good history since they do not reflect the evidence we have for the use of the Roman numerals. One could use the Roman numerical notation system to do calculations, just as one could use this a copy of this dissertation to protect oneself in a hailstorm – it is simply not a very efficient way to do things.

The best way to counteract the denigration of the Roman numerals is not, as Krenkel, Anderson, Detlefsen *et al.*, and others have done, to show that they are mildly (or even greatly) useful for a function for which they were never intended or used. Instead, we should examine the functions for which the system actually was or is intended and used (such as enumerating the above flaws), and evaluate its efficiency on that basis alone. The issue is not simply a game in the minds of mathematicians to see whether a system can serve some arithmetical function. By understanding the functions for which the Roman numerals were *not* used, we may be better able to understand the issues surrounding their eventual replacement.

Medieval - Additive

After the fall of the western Roman Empire, knowledge was distributed rather sparsely – for instance, among Western monks, Byzantine bureaucrats, and Arab scholars. The great early medieval Mediterranean polities – the Byzantine Empire and the early Muslim caliphates – mainly used ciphered-additive numerical notation systems such as the Greek and Arabic alphabetic numerals (ch. 5), and, later, ciphered-positional systems ancestral to the modern Western system (ch. 6). In Western Europe during the Middle Ages, the Roman numerals were the only ones in common use. Knowledge of other systems was restricted to peripheral regions such as Spain and southern Italy, and to a tiny well-educated elite. Even at the height of the Carolingian Renaissance (around

800 AD), arithmetic was essentially an art for the most learned members of society, and was not learned until quite late in the education of the early medieval scholar (Murray 1978). Still, while the need for large-scale bureaucracy and the corresponding need to express large numbers had declined since the height of the Roman Empire, Roman numerals were still frequently encountered, and even expanded in the range of functions they served.

The numeral-signs used in the early Middle Ages are very similar to those of the classical period. One difference was that rather than being written solely in majuscule characters, Roman numerals were occasionally written in the lower-case uncial script, although as Bischoff (1990: 176) points out, the majuscule numeral-signs predominated throughout the Middle Ages. When written in minuscule form, placing a stroke through the last numeral-sign of a numeral-phrase indicated that one-half was to be subtracted from the represented value. The medieval period also saw the introduction of a measure against fraud in numeral representation, namely that the last i in a numeral-phrase was often extended into a j, preventing anyone from adding additional signs to the end of the phrase (Menninger 1969: 285).

A more significant change was that alphabetic forms of the numerals for 500 (D) and 1000 (M) replaced the earlier Ɵ and ∞. This final assimilation of all the numeral-signs to alphabetic forms was the end-point of a long process of alphabetisation that had started in the late Republic. It would have had some mnemonic convenience; even if most of the numeral letters did not have any correspondence to the numbers they represented, at least literate learners of the system would not need to learn an entirely new set of signs. Alphabetisation of the Roman numeral-signs also enabled one to use them for numerical riddles, particularly *chronograms*, in which the values of the Roman numerals in a line of verse expressed the date of an event described in that verse (Menninger 1969: 281).

Structurally, the medieval system was largely unchanged from its classical predecessor. The multiplicative *vinculum* bar for 1000 used in classical antiquity continued to be used under a new name, the *titulus*, but the three-sided box symbol for multiplying by 100,000 was no longer used (Menninger 1969: 281). Large numbers were very rarely needed, and even the need for the *titulus* was limited. Subtractive forms also become more usual in the Middle Ages, though purely additive forms (e.g. IIII) were still commonly used.

A well-reasoned, if somewhat dated, paper by Shipley (1902) suggests that the changing use of Roman numerals between classical antiquity and the 9th century AD resulted in a number of regular errors in certain medieval manuscripts. In particular, by comparing the 5th century AD *Codex Puteanus*, containing sections of the works of Livy, and the 9th century AD *Codex Reginensis*, a copy of the former, the analysis of copying errors tells us much about the normal numeral-signs and structures used at the time of copying. Where the classical Roman form for 1000 was ∞, the medieval scribe was more accustomed to using M, and thus ∞ was often transcribed as X. Where the classical manuscripts contained Ɵ for 500, medieval scribes used D, and thus often omitted the Ɵ symbols entirely on the theory that the horizontal stroke indicated that the scribe had crossed out an error. Finally, because the subtractive form XL for 40 was used in medieval times as opposed to XXXX, instances of XXXX were abbreviated to XXX to correspond with correct medieval numeral-phrases. Shipley's analysis confirms the increased use of both subtractive structuring and alphabetic Roman numeral-signs as early as the 9th century AD.

The range of functions for which Roman numerals were used expanded considerably in the Middle Ages. Astronomical texts, which in antiquity were almost exclusively written with Greek numerals, often employed Roman numerals in the medieval era. As mentioned above, the alphabetizing of the signs for 500 and 1000 to D and M allowed the creation of number-riddles such as chronograms in which the numerical value of the Roman numerals in a phrase expressed the date of an event. Evidence for their use in legal documents and account-records increases greatly, though this may be a function of the differential survival of perishable materials from later periods.

Modifications and Replacement

The replacement of the Roman numerals by the Western system was not an easy and uncontested one. Instead, from the twelfth century until the seventeenth, there was great controversy throughout Western Europe regarding which system to use, with cultural, sociopolitical, and practical considerations being invoked in favour of one system or the other. Nor was the situation simply a choice between two options. In the majority of instances, one system or the other was adopted, with the Western system slowly coming to be accepted in the great majority of contexts. However, in a number of medieval documents, as knowledge of the ciphered-positional system spread into Western Europe, individual writers made idiosyncratic modifications to the Roman numerals in response to the interloping newcomer. While none of these modifications was adopted on a wider scale, their relevance lies in what they can tell us about the circumstances under which the Roman numerals were replaced by the Western numerals.

The most complete positionalized version of the Roman numerals is one of the earliest. Around 1130, the mathematician Ocreatus described a system using the Roman numeral-phrases for 1 through 9 (I, II, ... IX) in various positions as well as a special sign (O or t), which he called *teca* or *tsiphra*, to indicate an empty position (Smith and Karpinski 1911: 55; Murray 1978: 167). Positions were separated using a dot to avoid confusion. Thus, 1089 was expressed as I.O.VIII.IX. This system, which is obviously a blend of the Roman cumulative-additive and Arabic ciphered-positional systems, is cumulative-positional and has a mixed base of 5 and 10. While Murray characterises

Ocreatus' system as clumsy, it should be noted that it is fully positional, far more so than later compromises made between the Roman and Arabic systems.

While the use of barred numerals to indicate multiplication by 1000 was of early origin and continued throughout the Middle Ages, new multiplicative forms began to be used starting in the 13th century. Some examples of such numeral-phrases are listed below (Table 4.7), along with the transcriptions of the appropriate number in both the classical Roman system (including the use of subtractive and multiplicative forms, where appropriate) and in the Western system.

Table 4.7: Medieval multiplicative Roman numerals

Year	Numeral-Phrase	Source	Roman with multiplication	Western
1220	II•DCCC•XIIII	Menninger 1969: 285	$\overline{\mathrm{II}}\mathrm{DCCCXIV}$	2814
13th cen.	IIII•milia•ccc•L•VI	Menninger 1969: 285	$\overline{\mathrm{IIII}}\mathrm{CCCLVI}$	4356
1388	IIIIxx et huit	Guitel 1975: 225	LXXXVIII	88
1392	M C IIII IIII LXXIII	Cajori 1928: 33	$\overline{\mathrm{IIII}}\mathrm{CDLXXVIII}$	4473
1392	M XX III C IIII III	Cajori 1928: 33	$\overline{\mathrm{III}}\mathrm{CLXXXIII}$	3183
c.1500	Cd	Menninger 1969: 287	CIV	104
1502	XV.Cet: II	Menninger 1969: 285	MDII	1502
1505	I•V^{c}•V	Menninger 1969: 285	MDV	1505
1514	$\frac{\mathrm{II}^{\mathrm{C}}}{\mathrm{IIII}^{\mathrm{C}}.\mathrm{LX}}$	Cajori 1928: 34	$\frac{\mathrm{CC}}{\mathrm{CDLX}}$	$\frac{200}{460}$
1550	CCCM	Menninger 1969: 285	$\overline{\mathrm{CCC}}$	300000
1554	vi$^{\mathrm{M}}$ vii$^{\mathrm{C}}$ xiii	Menninger 1969: 283	$\overline{\mathrm{VI}}\mathrm{DCCXIII}$	6713
16th cen.	vij•c und XL	Cajori 1928: 34	DCCXL	740
1771	c m c i xxiij iiij lvj	Cajori 1928: 33	$\overline{\mathrm{CXXIII}}\mathrm{CDLVI}$	123456

These numeral-phrases primarily express multiplication by 100 or 1000 by juxtaposing C or M either immediately beside the appropriate multiplier, above it, or in superscript, and sometimes interposed with a dot. In two cases (the third and fifth in the above table), multiplication by 20 is expressed; the structure of these examples, which are from late medieval France, is a consequence of the assimilation of the system to the partly vigesimal structure of the French lexical numerals. All of the examples above are of a hybrid structure, purely cumulative-additive below a certain point and multiplicative-additive above. It might be thought that these multiplicative forms were adopted in order to write numerals more concisely or with a smaller set of numeral-signs - as was the case with the initial use of multiplicative forms in classical Rome. However, by comparing the numeral-phrases in Table 4.7 to their equivalents in standard Roman numerals, it can be seen that any such benefit is purely illusory. Regardless, in no way are they *positional* - the value of the numeral-signs does not change due to their position, but rather only due to their juxtaposition with another sign.

However, in addition to these multiplicative forms, we find many cases where the signs of the Roman system were intermingled with the positional principle of the newer Western system, as well as its actual numeral-signs. Menninger (1969: 287-8) provides examples of such admixtures starting in the late 15th century, as indicated in Table 4.8.

Table 4.8: Partially positional Roman numerals

M•CCCC•8II	1482
CC2	202
ICC00	1200
1•5•IIII	1504
15X5	1515
MDZ4	1624
MCCCC4XVII	1447
IV0II	1502

In these cases, where the system is no longer additive but begins to use position, it is obvious that conciseness is greatly increased. It is not altogether clear whether, in cases where the positional principle is used with Roman numeral-signs, this was being done consciously as a means of compromise between the two systems or as a misunderstanding of the Western system. Of course, such combinations are not necessarily advantageous; the idiosyncratic nature of these numeral formations almost certainly decreased their legibility. While they all use both the cumulative-additive and ciphered-positional principles, none does so in a way that can be clearly defined as a system. At any rate, by 1600, the Western numerals had achieved wide recognition and were well on their way to replacing the Roman system, and these hybrid formations no longer appear.

However, as I have already indicated, the replacement of the Roman numerals by Western ciphered-positional numerals was a drawn-out and highly contested process, lasting from around 1200 to 1600. While the Western numerals were first introduced to the West just before 1000 by Gerbert of Aurillac (later Pope Sylvester II), their use was infrequent before the publication of *Liber Abaci* by Leonardo of Pisa, also known as Fibonacci, in 1202. Fibonacci's mathematical text sparked an important debate between two camps: the *abacists*, those who preferred computation with the abacus and the corresponding Roman numerals, and the *algorithmicists*, whose pen-and-paper calculations using the Western ciphered-positional numerals were contentious, to say the least. The history of this debate is well documented, as it involved many important commercial families, renowned mathematicians and clergymen, and even state authorities (Murray 1978: 163-175; Menninger 1969: 422-445). Issues of computational efficiency were often addressed; one 16th-century advocate of the Western numerals claimed they were six times faster than abacus-calculation (Murray 1978: 166). However, these sorts of comparisons were not between the Roman and Western numerals, but between two techniques of computation: pen-and-paper arithmetic versus abacus calculation. The use of these techniques was correlated with specific numerical notation systems but evaluations of computational efficiency require more information about how numbers were manipulated.

Yet the debate was not only about efficiency. In 1299, the City Council of Florence issued an ordinance prohibiting the use of Arabic numerals in account-books because of the possibility of forgery; a similar edict was issued by the mayor of Frankfurt in 1494 (Menninger 1969: 426-427). One can certainly imagine the consternation of merchants and bookkeepers upon learning that one can simply add zeroes to the end of a numeral *ad infinitum* to multiply its value by ten each time. The rarity of paper in the earlier Middle Ages may also have contributed to the continued use of the Roman numerals (Smith and Ginsburg 1937: 29). Until pen-and-paper calculation became feasible, the "Roman numeral-friendly" abacus was the computational technology of choice, whereas the switch to pen-and-paper made the conciseness of Western numerals more attractive. This attractiveness only increased with the introduction of printing presses, where long strings of movable-type Roman numerals would have been unwieldy in comparison to Western numerals. We should also recognize a considerable cultural tradition at work. The Western numerals may have been rejected in part because of practical considerations, but they were, after all, a foreign and newfangled invention which involved the unusual principle of positionality and which (because of their novelty) could be used to conceal information or deceive others.

Despite such resistance, the Western numerals had taken hold by around 1300 in the Italian city-states. Elsewhere in Western Europe, particularly in Germany and England, the Roman numerals predominated until the late fifteenth century. Jenkinson (1926) finds limited evidence for the use of Western numerals in English archives before 1500, and notes that Roman numerals were only forbidden from use in state accounts in the 19th century. Barradas de Carvalho (1957) analyzed 15th and 16th century texts from Portugal, demonstrating that the Roman numerals were replaced there around 1500.

However, by the seventeenth century, the battle was essentially finished, and the Roman numerals ceased to be used in most contexts.

Today, the Roman system still enjoys a vestige of its former frequency of use, though it is encountered only in highly formal contexts or ones in which two distinct enumerations are needed. Yet why, after over a millennium of essentially unchallenged use in Western Europe, should the Roman numerals have ceased to be used? The traditional answer given – that Roman numerals were inefficient and thus were replaced – is correct but incomplete. The Roman numerals were used alongside the Greek alphabetic numerals for well over a thousand years, despite the great increase in conciseness that could have been achieved by abandoning the Roman system. In medieval Western Europe, the debate between the abacists and algorithmicists lasted into the 17th century, with the relative usefulness of the Roman numerals for calculation being greatly increased by the accompanying use of counting-boards. In particular, however, two developments appear to me to be fundamental in rendering the Roman system obsolete in the West. Firstly, the development of the printing press and the consequent rise in literacy after 1450 correlates very well with its rapid adoption throughout Europe (particularly northern regions). I believe this to be a consequence of the newly literate middle classes of Western Europe learning to read and calculate, unconstrained by centuries of traditional use of the Roman numerals. At the same time, the rise of mercantile capitalism and modern mathematics in the Renaissance changed the way in which numbers were viewed and used. While the subject of the connection between mathematics and capitalism is beyond the scope of this study, it is clear that the functional needs of Western society to represent number changed dramatically between 1300 and 1700, and that the computational functions that are better served by Western than by Roman numerals increased in importance.

Even today, the Roman numerals have not disappeared. The use of the Roman system on clock faces, in the enumeration of kings and popes, on many dated inscriptions, and copyright dates on films may be an archaic holdover from an earlier age, but it continues to hold an important place in the belief systems of many, through its symbolic connotations of antiquity, tradition, and prestige. Furthermore, in at least one of its remaining contexts of use - the pagination of introductions of books - it can be safely asserted to serve the very practical function of distinguishing introductory material from the body of the text. I do not believe that the retention of the Roman numerals in the future is guaranteed, nor nearly so likely as the continued use of Western numerals worldwide. I concur with the vast majority of researchers in regarding its value for performing many functions as minimal, and I would not expect any industrial society to be able to function solely with a cumulative-additive system such as the Roman numerals. However, the failure of over five hundred years' worth of predictions of its imminent demise suggests that these functional considerations, while important, do not tell the entire tale with respect to the history of numerical notation systems.

Arabico-Hispanic Variants

A number of important developments concerning the Roman numerals occurred in the Iberian peninsula between the tenth and sixteenth centuries - the period of the Reconquista and somewhat beyond, when Arabic and Western European knowledge systems interacted intimately for several centuries. The Roman numerals were well known in Spain and Portugal, as in the rest of Europe. The Arabic script had, since around 800 AD, used a ciphered-positional system much like the one used today, and of course also similar to the modern Western numerals. Medieval Spain presents us with a remarkable case where three separate systems coexisted, each using a different structuring principle: the cumulative-additive Roman numeral system and variations thereupon; the ciphered-additive Arabic alphabetic (*abjad*) numerals based on the Greek model; and the ciphered-positional Arabic numerals that are ancestral to those used today (cf. Labarta and Barceló 1988). But rather than a simple case of the two Arabic systems replacing the Roman numerals, we find a great deal of experimentation and admixture of systems on the part of medieval Spanish astronomers, bookkeepers, and scribes. In these systems, ciphering and positionality are incorporated in very innovative ways into the basic structure of the Roman numerals.

In the 10th and 11th centuries AD, Spanish astronomers were performing extensive astronomical calculations without the aid of any numerical notation system other than the Roman numerals. Even the use of the subtractive principle was very infrequent in Spain at this time. Despite the use of Arabic positional numerals in the region for some time previous and familiarity with works such as the *Arithmetic* of Al-Khwarizimi, Spanish astronomers did not adopt the system directly, despite its computational advantages (Lemay 1977: 458). Instead, several modifications were made to the Roman numerals to increase their compactness. Where the cumulative principle was normally used to express the numbers 3-9 (III, IIII, V, etc.), new acrophonic symbols based on the Latin numeral words were introduced: 3=t=*tres*; 4=q=*quatuor*; 5=Q=*quinque*; 6=s=*sex*; 7=S=*septem*; 8=o=*octo*; 9=N=*novem*. Of these, only the symbols for 4, 8, and 9 were commonly used, probably because they are the longest numeral-phrases below 10 in Roman numerals (IIII, VIII, VIIII). In addition, a special sign for 40 was used: X, a cursive ligatured version of XL, the only exception to the abandonment of subtractive forms (Lemay 1977: 459). Bischoff (1990: 176) reports that this sign for 40 is also found in some Visigothic manuscripts, suggesting considerable antiquity for the sign in the Iberian peninsula. These changes altered the classical Roman numerals into a partly ciphered-additive system. However, these modifications were not accompanied by the adoption of positional notation, though the acrophonic signs for 3-9 were probably indirectly stimulated by the Arabic notation; the system still had a mixed base of 5 and 10 and was purely additive. Regardless of their source, these developments increased the conciseness of numeral-phrases considerably. 99, which would have been LXXXXVIIII, could now be expressed as LXN. This system was used in various mathematical and astronomical texts until about the mid-twelfth century, at which time the Western numerals took hold in astronomy, as in the rest of Western Europe.

Yet while the Arabic positional numerals had firmly established themselves and later transformed into the Western numeral-signs familiar to us, the Roman numerals did not cease to be used in the Iberian peninsula, though their use became increasingly limited. Labarta and Barceló discuss two curious offshoot Roman numeral systems found in Spanish documents, as shown in Table 4.9 (Labarta and Barceló 1988: 32-34).

Table 4.9: Arabico-Hispanic numerals

1	5	10	50	100
ϭ	ح	ع	ســ	لس
I	Ↄ	O		

Both systems are purely cumulative-additive, decimal systems with a sub-base of 5, unlike those of the earlier astronomical texts, but use unusual numeral-signs. Unlike most Roman numerals, numeral-phrases are written from right to left (highest values at the right), which is curious because even in the Arabic script, which has a right-to-left direction, the numerical notation system has a left-to-right direction. The first of the above systems is found in a few late-16th and early-17th century documents to indicate monetary quantities. The numeral-signs are similar to letters of the Arabic script, though I am not qualified to evaluate whether or not they are acrophonic. The second system is obviously of very limited scope, and in fact is found only in a single Inquisition document from 1576 (Labarta and Barceló 1988: 34). Its similarity to the Berber numerical notation system (see below) suggests that the latter may have been derived from it.

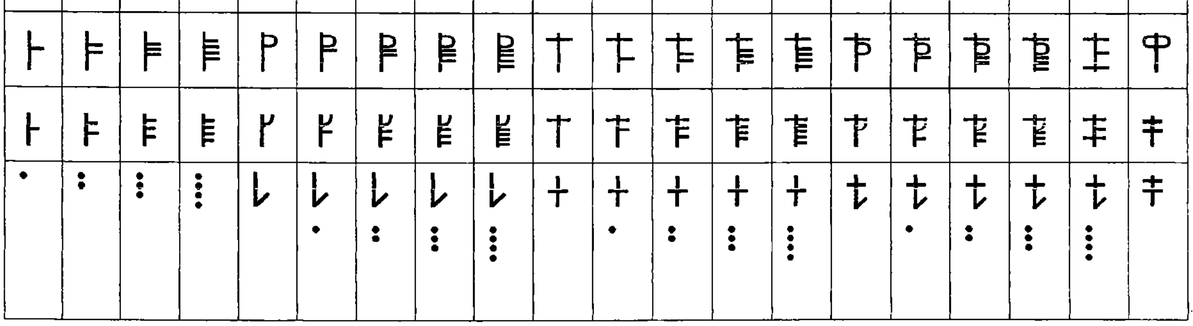
All the variant Roman numerals used in Iberian texts are relatively obscure and rare. In general, the standard Roman, Arabic, and Western systems were used, with the Roman numerals employed increasingly infrequently after the medieval period. By the 16th century, the Western numerals had established a firm foothold in most parts of the Iberian peninsula.

Calendar Numerals

In the 14th to 16th centuries, unusual numerals were used in certain documents and inscriptions from northern Europe pertaining especially to calendrical calculations. Known as "calendar numerals", "runic numerals", or "peasant numerals", they are clearly derived from the Roman numerals. I reject the term "runic" because most of the numerals do not occur in runic inscriptions and because a different and unrelated runic numbering system was sometimes used. Likewise, the term "peasant numerals" tells us who may (or may not) have been using them, but lacks the precise functional association of "calendar numerals". Some examples of these signs are shown in Table 4.10 (cf. Ifrah 1985: 146-147; Kroman 1974: 121).

Table 4.10: Calendar numerals

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20



These systems are all cumulative-additive and have a base of 10 with a sub-base of 5. Units are marked by strokes or dots, fives are marked by angled or curved lines or loops to create "U" or "V" shapes, and tens are marked by transecting the vertical line perpendicularly, creating a cross or X. Although these numeral-signs are often joined together into single figures resembling digits by using a vertical stroke, the system is not a ciphered one. There is no evidence of signs for 50, 100 or higher values in these systems, and I am not aware of calendar numerals being used for numbers higher than around 30. Because of the specialized function of these numerals, there was almost never any reason to express numbers higher than 19 using this system.

Other than their unusual numeral-signs, the calendar numerals are identical to ordinary Roman numerals. I do not find the classification of these numerals as "tallies" to be particularly useful (Menninger 1969: 249-251; Ifrah 1985: 146-147). Both Menninger and Ifrah regard calendar numerals as part of a cultural substratum of tallying and notching leading back into the depths of prehistory. I do not deny that tallies are very ancient, or that they are distinct from numerical notation systems; rather, I deny that the calendar-numerals are tallies. A tally is used ordinally - one places marks as necessary on some material (wood, paper, stone, etc.) in sequence in order to keep a running total, rather than, as with numerical notation systems, marking an already-totalled value. Simply marking Roman numerals in a slightly unusual way (in this case, vertically, attaching the numeral-signs to a vertical line) does not make the system any less an offshoot of the Roman numerals, or any less a numerical notation system.

By the 15th century, the Western numerals were fairly well known in Germany and were becoming much more common in England and Scandinavia. Furthermore, all three regions were using the Roman numerals for various purposes at this time. Calendar-numerals were used in a very delimited set of contexts, namely on documents of wood, stone, or horn designed to assist the largely illiterate populace in determining the dates of festivals, especially Easter. Throughout the medieval period, the Metonic cycle of 19 years, after which the moon's phases recur on the same day of the year, was used as a rough-and-ready guide to calculate the date of Easter. The function of the calendar-numerals was solely to denote the various years of this cycle on stylized perpetual calendars, known as "runic calendars" in German and Norse-speaking areas and as "clog almanacs" in England. While these texts were used for computation, the numerals were not used directly for arithmetic. One merely needed to line up the appropriate days and years to get the correct value. Calendar numerals were, however, reasonably compact and easily understandable by anyone who knew the Roman numerals.

The calendar numerals lie at the heart of one of the major pseudo-scientific controversies of New World archaeology – the so-called "Kensington Rune Stone" of Minnesota, which purportedly contains a genuine Viking inscription left there in 1362 by Norse explorers who had travelled westwards from Vinland. At issue are the calendar or runic numerals used in the inscription, particularly the numbers 14, 22, and 1362, expressed thereupon as ├ ≣, ╞ ╞, and ├ ≡ ᚹ ╞, respectively (Nielsen 1986: 51). Clearly, these numerals are structured in a ciphered-positional fashion, like our Western numerals but unlike the cumulative-additive calendar numerals. Struik points out the improbability that a Norse explorer in "Vinland" would have been familiar with the Western numerals in 1362, which were only known by the very educated in northern Europe at the time (Struik 1964: 167). Even if this knowledge is presumed, we need to explain why the savant substituted the cumulative-additive runic numeral-signs for the appropriate Western figures. Additionally, I know of no evidence for the use of calendar numerals except in calendrical texts, which the Kensington stone clearly is not. Nielsen (1986) has feebly attempted to counter this argument by showing that some runic inscriptions used Western numerals at this time, but even if the dating of his evidence is correct, the numerals he notes are still recognizably Western numerals, not calendar numerals. It is far more likely that Ohman, the Swedish-American "discoverer" of the stone, was familiar with the runic numeral-signs (easily found in any book on runology) but not with the proper (cumulative-additive) structure of the system.

The calendar numerals are thus an interesting sidebar to the general history of the Roman numerals. Essentially a local phenomenon in northern Europe, their passing was largely due to the increasing use of Western numerals for calendrical and other computational purposes throughout the Renaissance and early modern periods. By 1643, when Ole Worm wrote his *Fasti Danici* describing the "runic" numerals used in calendar tables, it was primarily as a curiosity and to aid the transition to newer methods of calculation (Worm 1643).

Greek acrophonic

During the period from roughly 750 to 500 BC, what we now call archaic Greece was a conglomeration of generally monarchical and invariably small city-states in mainland Greece, the Aegean islands (including Crete), the southern half of the Italic peninsula (known as Magna Graecia) and western Asia Minor, sharing in common only the use of Greek dialects. A tremendous number of local varieties of writing, known as *epichoric* scripts (from Greek *epi-*, upon, over, and *chora*, place, country) were used during this period. These scripts, from which the modern Greek one is descended, were developed on the model of the Phoenician consonantary no later than 800 BC. In their earliest phases, some of these alphabets were written from right to left or in alternating directions (*boustrophedon*), although by around 500 BC all the epichoric scripts were written from left to right. Adjoining these scripts were two very distinct types of numerical notation: the acrophonic, described below, and the ciphered-additive alphabetic numerals (ch. 5). For our present state of knowledge of these two systems, we are greatly indebted to the tireless work of Marcus Niebuhr Tod, whose research on the acrophonic system particularly is unparalleled.[3]

The Greek acrophonic system is so named because the signs for many numbers are taken from the first letter (*akros* = highest, outermost; *phone* = sound) of the corresponding (classical) Greek word: Γ = ΠΕΝΤΕ = 5; Δ = ΔΕΚΑ = 10; Η = HEKATON = 100; Χ = ΧΙΛΙΟΣ = 1000; Μ = ΜΥΡΙΟΣ = 10000. The system is cumulative-additive, uses vertical strokes for units, has a base of 10 with a sub-base of 5, and is always written

[3] Tod's six papers on Greek numerical notation (Tod 1911-2, 1913, 1926-7, 1936-7, 1950, 1954) have been reprinted in one volume (Tod 1979). My citations are taken from the original papers themselves.

from left to right, with numeral-phrases in descending order of numeral-sign value. The signs for 50, 500, 5000, and 50,000 are represented by combining the sign for 5 with the sign for the appropriate multiplier. Whether we choose to see these secondary base numeral-signs as single signs or as two ligatured ones using the multiplicative principle is largely a matter of definition, and does not substantially affect how we classify the entire system. A great deal of unnecessary verbiage has been devoted to the nomenclature of this system (Tod 1911-12: 125-127). It is true that the secondary signs (multiples of 5) are not purely acrophonic and the sign for 1 is simply a vertical stroke. However, there is no reason to think that any of the other names (Herodianic, decimal, etc.) proposed for the system are any more accurate for the purposes of historical analysis than the widely preferred designation "acrophonic". The acrophonic system as used in classical Athens is shown in Table 4.11.

Table 4.11: Greek acrophonic numerals (Tod 1911-12: 100-101)

1	5	10	50	100	500	1000	5000	10000	50000
I	Γ	Δ	𐅄	H	𐅅	X	𐅆	M	𐅇

Thus, 36849 would be expressed as MMM𐅆X𐅅HHHΔΔΔΓIIII. This particular set of acrophonic signs was used at Athens throughout the system's history; similar acrophonic signs were used in large portions of the Hellenic world, the only difference being that the appropriate letters from each epichoric script were used in place of the letters used in the Attic inscriptions. Dow (1952) notes that the variety of acrophonic Greek numerical notation systems stands in sharp contrast to the Greek alphabetic system, which is remarkably consistent throughout its geographic and temporal range. Ironically, the degree of variation among local systems is far greater than the variety of lexical numerals used in the Greek dialects, thus negating its cross-cultural translatability, a major advantage of numerical notation. This has led some to raise the issue that the general paucity of acrophonic numerals may result partly from their incomprehensibility in international commerce, though whether they would in fact have been incomprehensible is doubtful (cf. Tod 1936-7: 246).

For expressing monetary values, the acrophonic numerals of various regions could be modified to reflect the forms of currency being expressed. These contexts are also the only ones in which fractional values occur. For instance, Threatte describes the following symbols used for Athenian currency: T (talanton = 1 talent = 6000 drachmas), M (mna = 1 mina = 100 drachmas), Σ (1 stater), ├ (1 drachma), I (1 obol), C (1/2 obol), Ɔ or T (1/4 obol), and X (1/8 obol) (Threatte 1980: 111). For the talent, mina, and stater, multiplicative or ligatured numeral-signs were sometimes used to express a value. While there is some potential for confusion in this system (T can mean 1 talent or ¼ obol; M can mean 1 mina or the numeral 10,000, etc.), most problems are avoided by the fact that numeral-signs are always listed in descending order. In some regions, special signs were used to indicate monetary values that did not fit easily into the standard system. For instance, a system is found in inscriptions from Thespiae (in Boeotia) that uses numeral-signs for 30 and 300, which consist of a sign T (for *triobole*, or 3 obols) ligatured to the appropriate sign for 10 or 100 (Tod 1911-12: 109; Feyel 1937).

Despite the name of the system, not all numeral-signs used in the Greek epichoric scripts are acrophonic. Johnston (1975, 1979, 1982) has found several instances of a very early Greek cumulative-additive but non-acrophonic system with a mixed base of 5 and 10 dating from the 6th and 5th centuries BC and found throughout the Greek world. The signs of the system are shown in Table 4.12 (cf. Johnston 1979: 29-30; Johnston 1982: 208).

Table 4.12: "Non-acrophonic" Greek numerals

1	5	10	50	100
I	Γ	↑	ᛍ	ᛡ

Johnston argues that this system was built up systematically by cumulatively adding oblique lines to a vertical stroke to obtain higher numeral-signs. Curiously, he does not note that the signs for the sub-base (5 and 50) are the right halves of the appropriate primary bases (10 and 100). Here we see a clear parallel to the halving of Etruscan numeral-signs, which is notable because many of the examples of this "pre-acrophonic" system are of South Italian provenance.

A very unusual numerical notation system used only to express monetary values is found in five 4th century BC Greek inscriptions from Cyrene (in modern Libya), which was a Greek colony for several centuries. The numerals found there are non-acrophonic and their interpretation has been a matter of controversy for many decades (Tod 1926-7; Oliverio 1933; Tod 1936-7; Gasperini 1986). Our best evidence comes from the temple of Demeter at Cyrene, where inscriptions list the prices of various goods and the temple's revenues and expenditures (Tod 1936-7: 255). They are particularly odd in that they present a dual series of figures in which each numeral-sign has both a higher and lower value; the specific amount must be inferred from the context within the numeral-phrase. The interpretation presented by Oliverio, Tod, and Gasperini is derived from an analysis of the maximum number of times each sign is repeated (and is thus open to question if more inscriptions are found). The signs with their values under this interpretation are shown in Table 4.13.[4]

Table 4.13: Cyrenaic numerals

20000	10000	5000	1000	500	100	20	4	2	1	1/5	1/10	1/50
⋝	⊏	Z	>	Γ	—	X	⋝	⊏	Z	>	—	Δ

The majority of Cyrenaic numeral-signs have two values, of which the higher is 5000 times the value of the lower. I am unable to explain why this breaks down for some of the lower signs. Most of the difficulties in explaining the signs used in this system are due to the relative values of different units of currency used in Cyrene during this period

[4] The numbers listed are amounts in drachmas, based on the assumption that the lower Z sign represents one drachma, without which the absolute value of each sign would be indeterminate.

(drachmas, staters, minas, and talents, where 1 talent = 50 minas, 1250 staters, or 5000 drachmas). The function for which this system was used is no help in explaining its irregularity, as temple records are a common function for acrophonic numerals throughout the Greek world. I suspect that the unusual nature of the system is due to a local metrological or monetary system whose nature is not clear. Because we do not have abundant acrophonic numerals from other contexts in Cyrene, we do not know whether this system was employed for non-monetary functions.

Perhaps the most interesting developments in the acrophonic system are found in 4th century BC inscriptions, not from peripheral areas but from the core of Greek culture. In Olynthus (in the northern Chalcidice region), a numerical notation system was used which is non-acrophonic and which lacks a sub-base of 5 (Tod 1936-7: 248-9; Graham 1969). This system is interesting both because the signs for 10, 100, and 1000 (X, 8, and Ψ, respectively) are the last three letters of the western Greek alphabet used in the region, and because the sign X = 10 is common to Roman and Etruscan inscriptions as well.[5] On this basis, Graham (1969: 356) argues that the Roman/Etruscan system was borrowed from the Chalcidian colony at Cumae (in southern Italy). This theory, while attractive, has several flaws, many of which derive from Mommsen's (1965 [1909]) "lost-letter" theory of the Roman numerals, which I rejected earlier. The most serious problem I perceive with Graham's theory is that the 4th century BC numeral-signs of Olynthus are supposed to have spread to the 6th century BC Etruscans by means of a colony at Cumae that never used the numeral-signs in question. Regardless, the Olynthian numerals are very intriguing, and their relationship to the Greek alphabet seems clear (though they are just as clearly *not* acrophonic). I suspect that the letters were borrowed for the higher exponents just as the Romans began with non-alphabetic numeral-signs, but later modified their signs into alphabetic ones for mnemonic purposes.

---

[5] No significance should be attributed to the fact that the sign ∞, a common Roman numeral-sign for 1000, is rotated 90° from the Olynthian sign 8 for 100.

A similar system was used in Epidaurus, on the southern Greek mainland (Tod 1911-12: 103-4). It is acrophonic for 100 and 1000 but not for the lower exponents. Nearby, in Argos and Nemea, a closely related system was used that apparently had a sign for 50, but not for 5 (Tod 1911-12: 102-3; Ifrah 1985: 235). The systems of Epidaurus and Argos are unique among the Italic numerical notation systems in their use of ● rather than I as the sign for 1. Very few inscriptions from this region contain numerals, and the cultural history of these systems is in serious need of analysis. Table 4.14 summarizes the numeral-signs used in these irregular systems as compared to the standard acrophonic system:

Table 4.14: Epichoric Greek numerals

System	1	5	10	50	100	500	1000
Standard Acrophonic	I	Γ	Δ	𐅄	H	𐅅	X
Olynthus	I	None	X	None	8	None	Ψ
Epidaurus	●	None	—	None	⊟	None	X
Argos and Nemea	●	None	O ⊙	𐅄 Γ	⊟	None	X

The origins of the standard acrophonic system remain obscure. None of the forms of the number-signs, save the vertical stroke for the units, has any relation to any other numerical notation system. While this is to be expected, given the use of the acrophonic principle, it makes reconstructing the history of the system rather difficult. It appears to be the implicit assumption of most classicists that the system was invented independently from the Roman, Phoenician, and other systems used at the time. Ste. Croix (1956:52) makes this statement explicitly, but gives no justification for it. It could be argued that the acrophonic nature of the system suggests that it could only have been invented in Greece. However, I have already shown that not all the systems in use in Greece were acrophonic. It is very likely that the earliest "acrophonic" systems were not acrophonic at all, and that only later were numeral-signs assimilated to alphabetic forms.

To examine this point further, we require a crucial datum – the temporal and geographical context of the system's invention.

The traditional dates given for the use of the acrophonic system in Athens are 454 to 95 BC, and these figures have been widely quoted in modern histories of mathematics as encompassing the entire duration of the system (Heath 1921: 30). Yet there is considerable evidence of a much earlier origin for the acrophonic numerals. Tod argues, solely on logical grounds, that a 7th century BC origin is not unreasonable, given that the system was fully developed by the middle of the 5th century BC (Tod 1911-12: 128). Mabel Lang mentions a 7th century BC decorated Greek amphora inscribed with the number three as **III**, but this certainly does not imply that the numeral in question was part of the acrophonic system; it simply might have been part of an unstructured tallying-system or almost any other system in use in the Aegean at the time (Lang 1956: 3). All of Lang's other examples of numerals on vases are from the 5th century BC or later.

For the second half of the 6th century BC, however, there is more promising evidence of the acrophonic system. Johnston (1979: 27-29) discusses three different variations of the 'pre-acrophonic' system mentioned above, which differ mostly on palaeographic grounds. He finds evidence of 6th century BC use of these signs in southern Italy, Sicily, western Asia Minor, the Aegean islands, and various parts of mainland Greece – in short, almost the entirety of Greek civilization during that period. Several vases from southern Italy and Sicily, which Johnston dates to the last quarter of the 6th century BC, bear marks used in commercial transactions (Johnston 1975, 1979, 1982). In particular, I find it telling that so many variations of the acrophonic system are known in the 5th and 4th centuries BC, with increasing regularity emerging as the system develops. This suggests an initial period of experimentation followed by consolidation and agreement on a single form of the numerals.

I believe it entirely possible, though not proven, that the Greek acrophonic numerals ultimately originated on the Italian peninsula around 575 - 550 BC, either independently or on the model of the Etruscan system. As I have provisionally accepted Keyser's (1988) contention that the Etruscan numerals developed relatively independently as an outgrowth of tally-marks, the obvious conclusion is that the Greek system developed on the model of the Etruscan numerals in southern Italy and Sicily, an area of considerable commercial and cultural contact between the two cultures. I find it difficult to believe that two cumulative-additive, quinary/decimal numerical notation systems developed on the Italian peninsula in the second half of the 6th century BC independently of one another. Obviously, the situation is complex; the Etruscans owe their script and many other features of their culture to contact with the Greeks; it is not usual to think of the transmission of ideas moving in the opposite direction, and indeed it remains possible that the Etruscan numerals have a Greek origin. We must also deal with the possible influence of the Phoenician colonies in North Africa and western Sicily, which were in contact with both groups during the 6th century BC. More examples are needed before any firm conclusions can be reached.

I have already detailed the diffusion of the acrophonic system throughout the Greek-speaking world. In the early classical period, acrophonic numerals were used in Asia Minor, the Aegean islands, North Africa, southern Italy, and Sicily, in addition to mainland Greece. Yet its spread to the non-Greek world was relatively limited. The Lycians of southern Asia Minor used a non-acrophonic numerical notation system in the late 5th and 4th centuries BC that is probably an epichoric variant of the acrophonic system even though their language was not Greek (see below). The enormous cultural debt of Lycia to classical Greece is beyond doubt, and its geographic and temporal proximity strengthens this hypothesis. More speculative, but surely possible, is the possibility that the South Arabian numerals, which arose in the 5th century BC, derive from the acrophonic system. The South Arabian numerals are cumulative-additive, are base-10 with a sub-base of 5, and use acrophonic numeral-signs, and I believe their origin to be

Greek. However, more evidence of cultural contact is needed before such a hypothesis can be proven.

Throughout its history, the acrophonic numerals were used for a surprisingly small number of functions. They are found on inscriptions on stone, lead, and silver as well as on potsherds; they may have also been used on wood or other perishable materials, though evidence is lacking. Of the thousands of Greek papyri from the 4th century BC onwards, only a handful from Saqqara contain acrophonic numerals (Turner 1975). Inscriptions on stone containing numerals include accounts, inventories, lists, regulations, treaties, and boundary-markers, as well as graffiti or other marks on pottery to indicate quantities for commercial purposes. The acrophonic numerals expressed measures of volume or distance, quantities of goods, or monetary values. As mentioned earlier, the numeral-signs differed somewhat when used for the last of these purposes.

What is notable is the wide range of purposes for which acrophonic numerals were *not* used, even compared to other cumulative-additive systems used in the Mediterranean in antiquity. Firstly, the numerals could only be used to express cardinal numbers; ordinal numbers were expressed either through lexical number words or, when available, by using alphabetic numerals (Tod 1911: 128). The Greeks never expressed dates of any kind in acrophonic numerals, as they did not use a standardized dating system, except in some later inscriptions where regnal years were expressed lexically or through alphabetic numerals. The practice of expressing the age of the deceased at death on funerary inscriptions, a source of much information on other numerical notation systems, does not seem to have been the custom in Greece. Threatte notes that documents in connected prose (decrees, for instance) do not normally contain acrophonic numerals, except to indicate the price of executing the inscription (Threatte 1980: 112).

There is no evidence that the acrophonic numerals were used direcly for arithmetic or accounting. For these purposes, as with the Roman and Etruscan systems, the Greek acrophonic system was supplemented by the use of the pebble-board style abacus, in which several grooves were labelled with the appropriate acrophonic numerals. Lang (1957) has established that many of the mathematical errors made by Herodotos demonstrate his use of the abacus do perform calculations, with which certain types of errors (especially in multiplication and division) can occur very easily. She also lists the thirteen known examples of abaci (and fragments thereof) known from classical Greece, all of which have a series of acrophonic numerals inscribed in a row (Lang 1957: 275-6). While this number may seem low, it is actually far more evidence than we have for the classical Roman use of the abacus. Most notable among these abaci is the remarkably well-preserved "Salamis tablet", which probably dates from the 5th century BC (Menninger 1969: 299-303). The numerals on it range from T (one talent) to X (1/8 obol); the monetary values of the numeral-signs suggest that it was used for practical commercial computations.

The decline of the acrophonic system is thoroughly entwined with the fate of the Athenian state as a Greek power. From its height in the inscriptions of the 4th century BC, it was slowly replaced as Athens ceased to be a dominant power in Mediterranean affairs. By the 3rd century BC, the acrophonic system had been largely replaced by the alphabetic numerals for most purposes throughout large parts of the Hellenistic world, including Ptolemaic Egypt and Seleucid Persia. Only in Athens and the surrounding areas did the acrophonic system continue to flourish. Threatte lists all known 1st century BC examples from Athens – only a handful (Threatte 1980: 113). By this time, Greece was of course firmly under Roman control. Yet there is no evidence that the acrophonic system was replaced by Roman numerals except, as one might expect, in southern Italy, as Latin-speaking populations came to dominate in that region. However, the use of acrophonic numerals did continue in one very limited domain – stichometry, or the enumeration of lines of verse in classical texts (Tod 1911-12: 129-30). This practice continued as late as the 3rd century AD with the writings of the Neoplatonist philosopher Iamblichus. Such late examples are analogous to the use of Roman numerals in contexts in the modern West, in which it is very useful to have two separate numerical notation systems - for paginating introductory sections versus the body of a work, or for distinguishing volume and page numbers of certain texts.

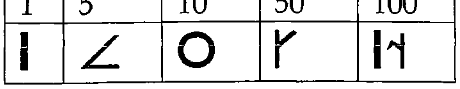
Lycian

Lycia was a small state of southern Asia Minor in the middle of the first millennium BC, centred around the city of Xanthus. The Lycians spoke an incompletely understood Indo-European language related to the earlier Luwian language, which was spoken in the Neo-Hittite kingdoms of Asia Minor until around 700 BC. Throughout its history, Lycia occupied an intermediate position between the Greek and Persian spheres of influence, and was intimately involved in interregional commerce and conflict. The Lycian alphabet, which was developed around 500 BC, is clearly an epichoric variant of the Greek script, like many others used in the Greek peninsula and western Asia Minor, with the only difference being that the language of the inscriptions was not a Greek dialect. A few hundred instances of the Lycian script have survived, mostly from inscriptions on stone and on coins; they are written almost exclusively from left to right and date to the 5th and 4th centuries BC.

The Lycian numerical notation system is still very poorly understood. It appears that the Lycian system, like the Greek acrophonic, Etruscan and Roman systems, is cumulative-additive and decimal in structure with a quinary sub-base. However, the exact values of the numeral-signs are still in debate, and I cannot hope to settle the matter finally, but merely to present the evidence we currently have. It is generally agreed that the values accepted by Shafer (1950) and Bryce (1976) for the lower numerals are correct. The signs of this system, under this interpretation, are shown in Table 4.15.

Table 4.15: Lycian numerals

1	5	10	50	100



Thus, 127 would be expressed as I˥OO∠II. There is also a sign, —, that probably represents ½, although Shafer (1950: 260) argues that it may represent an additional one-half of any numeral-sign that precedes it; O— would be 15 and ∠— would be 7½ according to this theory. There is some controversy with respect to the numeral-signs for 50 and 100, which are found only on a few inscriptions. The value 50 is assigned to Γ primarily by default; its value is certainly between 10 and 100 (it is found after I˥ but before O). I follow Frei (1976: 15) in assigning the value of 50. We can be fairly certain about the value of the sign for 100, because it is found in the Lycian portion of a trilingual Greek-Lycian-Aramaic inscription found at Letoon and dating from 358 BC (Frei 1976: 13-15). Shafer's contention that the Lycian signs for 10 and 100 are identical, that the Lycians possessed an unattested sign for 0 and that there are unique Lycian numeral-signs for 6 and 7 is quite bizarre (Shafer 1950: 261). Shafer's own doubts on the validity of the transcription seem quite warranted. I am unconvinced that such an irregular system ever existed, and the sign-values I have presented in Table 4.15 seem far more plausible.

Shafer (1950:258-9) suggests that the Lycians may have used the subtractive principle to express the number 4 as I∠, in particular because of an inscription in which a husband, his wife, his ∠ children (sons?) and their I∠ wives are buried. Shafer contends that since there is no evidence that the Lycians practiced polygyny, it is unlikely that five men would be buried with six wives, and so I∠ must mean 4. However, this is extremely thin evidence on which to postulate such a feature. One or more sons might have remarried after the death of a first wife. Moreover, in other inscriptions, 4 is expressed as IIII. Finally, there is no evidence of the use of subtractive forms by the Romans or anyone else at this early date. I am thus very dubious regarding the use of the subtractive principle in the Lycian numerals.

The origin of the Lycian numerals has not yet been firmly established. Because the Lycian numerals arise in the early 5th century BC at the time of the peak use of the

Greek acrophonic numerals in Athens and throughout the Hellenic world, the theory that the Lycian numerals derive from the Greek acrophonic numerals deserves serious consideration. Both systems were purely cumulative-additive, had a base of 10 with a sub-base of 5, and were used in the 5th century BC in the Aegean region. While the Lycian numerals are clearly not acrophonic, many of the epichoric numerical notation systems of the classical Greek world were not acrophonic, but used a wide variety of symbols for 5, 10, 50, and 100. I thus conclude that the Lycian numerals are probably a previously unidentified variant of the Greek cumulative-additive systems.

However, Frei's alternate hypothesis deserves some note, namely that the Lycian system was based on an Aramaic model, and that the sign for 100 ( ) is in fact multiplicative (1x100) rather than constituting a single numeral-sign (Frei 1976, 1977). This is a particularly relevant datum because the Lydian script, which was used in Asia Minor at the same time as the Lycian and which is closely related to it, did not use numerical notation based on Greek but rather used the Aramaic system unmodified. Because none of the Semitic systems had separate signs for 50, but all of them had signs for 20, we would need to modify the value of the Lycian to 20, which is consistent with the numeral-phrases known from inscriptions. However, three factors suggest that the theory of Aramaic or Phoenician origin is less parsimonious than that of diffusion from the Greek acrophonic system. Firstly, there is generally little similarity between the numeral-signs of Lycian and either Phoenician or Aramaic. Secondly, the Lycian system clearly uses a sign for 5, which is encountered only very rarely in Aramaic inscriptions from this date, and not at all in Phoenician inscriptions. Thirdly, Lycian, like the acrophonic numerals, is written from left to right, whereas the Levantine systems are all written from right to left. Yet because no numbers higher than 120 are expressed in any Lycian inscriptions, it cannot be established whether the Lycian numeral-phrase for 200 and higher multiples of 100 were additive ( ) or multiplicative ( ). Clearly, more evidence is needed before a final judgement on the issue can be provided.

The Lycian numerical notation system apparently did not diffuse outside Lycia. It is impossible that the Lycian numerals led to the development of the acrophonic system, as we have several 6th century BC examples of the latter system. Although Shafer (1950) argues that the similarities between the Roman and Lycian numerals are sufficient to indicate the derivation of the former from the latter, this likeness is no greater than between Lycian and the Greek acrophonic system. Furthermore, while there is some similarity between the Lycian numeral-signs and other systems of the Italic family (especially Berber and Minaeo-Sabaean), these similarities do not correspond to any plausible circumstances of cultural contact between these regions. In Asia Minor, scripts such as Phrygian and Lydian, both of which are closely related to Lycian and were used in the 5th and 4th centuries BC, used numerical notation based on the Phoenician-Aramaic model rather than on the Greek.

Lycian numerals are found primarily in a single context, on sepulchral epitaphs indicating monetary amounts, normally including a numeral-phrase preceded by the word *ada,* now considered to be a monetary unit (Bryce 1976: 175). It has traditionally been argued that the monetary values stipulated a penalty to be paid should the tomb in question be violated (Shafer 1950). More recently, Bryce has argued that this interpretation may be flawed and that the values indicate fees paid in advance by the family for a tomb site (Bryce 1976). The only non-funereal context where Lycian numerals are used is the trilingual inscription found at Letoon, which is generally regarded to be a public legal regulation (Frei 1976). Regardless, because Lycian numerals are not found on coins or on financial inscriptions, they are quite distinct from the numerals of the rest of Asia Minor and the Aegean.

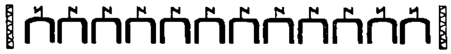
As the Lycians became increasingly caught up in imperial conflicts between the Persians and the Greeks (both classical and Macedonian), their script came to be used increasingly infrequently. By 300 BC, the Lycian script had assimilated to the Greek, and its numerical notation ceased to be used, replaced by the Greek alphabetic numerals.

South Arabian

The Old South Arabian scripts are of a very ancient origin, first appearing around the turn of the 1st millennium BC in the southern part of the Arabian peninsula (modern Yemen). They are consonantal and are characterized by large, well-formed letters and by their extremely varied direction of writing (left-to-right, right-to-left, or boustrophedon, depending on the inscription). They were used for well over a millennium to write South Arabian languages such as Minaean, Sabaean, Qatabanian, and Hadramauti. During their early history, these scripts did not possess any numerical notation system. At the time of the rise of the kingdoms of Minaea and Saba in the 5th century BC, numerical notation began to be used in South Arabian monumental inscriptions. The numeral-signs used are shown in Table 4.16, including both left-to-right and right-to-left sign forms, where appropriate (Hommel 1893: 8).

Table 4.16: South Arabian numerals

	1	5	10	50	100	1000
Left to right	𐩽	𐩭	𐩲	𐩾	𐩣	𐩱
Right to left	𐩽	𐩭	𐩲	𐩾	𐩣	𐩱

The system is cumulative-additive, with a base of 10 and a sub-base of 5, and is written in whichever script direction is used in the inscription as a whole. The sign for 1 is, as in all systems of the Italic family, purely iconic. The signs for 5, 10, 100, and 1000, however, are acrophonic; each is simply the first letter of the appropriate South Arabian lexical numeral (Beeston 1984: 8). The sign for 50 is non-acrophonic, but is simply a halved version of the sign for 100. In one inscription (Biella 1982: 531), the sign X is apparently used with the numerical value 4, possibly in imitation of the Nabataean system (ch. 3). There are no signs for 500 or 5000 known from any South Arabian inscriptions; in inscriptions, 500 was written as 𐩣𐩣𐩣𐩣𐩣 (Biella 1982: 265) and 5000 as 𐩱𐩱𐩱𐩱𐩱 (Biella 1982: 1). Normally, numeral-phrases were placed between large hatched bars () to avoid confusing numeral-phrases with words, given the use of the acrophonic principle (Halevy 1875: 78). Thus, 3697 could be expressed as  or as . Large sets of unit-signs were not divided into smaller groups, which presents a problem because the sub-base of 5 is not used throughout the system; one inscription lists 12,000 as  (Ifrah 1998: 187).

In some inscriptions, the South Arabian numerals used an unusual technique of implied multiplication that resembles positional notation. Most often, this was done when a value greater than 10,000 was expressed, by placing signs to the left of a sign for 1000 (when reading from left to right) which were implicitly taken to represent multiples of 1000. For instance, one inscription has  for 31,000, in which the 3  signs have the value of 10,000 instead of 10, while  retains its ordinary value of 1000 (Biella 1982: 349; Ifrah 1998: 187). Apparently this technique was also sometimes used for multiples of 100; Halevy (1875: 79) notes an inscription that has  instead of  for 300. We know that the multiplied value is correct because of contextual information and because South Arabian inscriptions commonly list the appropriate lexical numeral beside the numeral-phrase. Ifrah sees in this use of implied multiplication "what might be called the germ of our place-value notation" (Ifrah 1985: 232). However, without contextual information, such numeral-phrases would simply be confusing and ambiguous, as there is no sign for zero. At any rate, these formations are very rarely attested throughout the system's history.

Despite its unusual structural features, it is generally agreed that the South Arabian system is derived from the Greek acrophonic system (Ifrah 1998: 186; Fevrier 1948: 579). Both systems are decimal and cumulative-additive and both have a sub-base of 5. Additionally, both systems use the acrophonic principle, a feature that may also be present in the Berber system (see below), but is otherwise uncommon at this period. The numeral-signs themselves are not similar in form to any other system, but this is not too surprising, since the system is acrophonic. More notable is the fact that the South Arabian numerals are first encountered in the 5th century BC. The Minaeans and Sabaeans were actively engaged in trade with the Greeks at this time, when the acrophonic numerals were the only ones the Greeks were using for monetary and metrological purposes. While it might be expected that the South Arabian scripts would have used numerical notation systems similar to those used in North Semitic scripts at the time (Aramaic or Phoenician), this is not borne out by comparing the systems. The sign for 20 is absent in the South Arabian system, while the Aramaic and Phoenician systems did not normally use signs for 5 and 50.

There is some reason to believe that the South Arabian script and numerical notation system are ancestral to those used in North Africa from about 250 BC to 250 AD. Yet there are problems with establishing whether there was significant cultural contact between these two regions, which are separated geographically by over 3000 km. Other than this, however, there is no evidence that the South Arabian numerals ever diffused outside the Arabian peninsula. The Ge'ez script used for the Ethiopic languages, which is derived from a South Arabian model, used numerals based on the Greek alphabetic system.

Although some South Arabian cursive inscriptions on wood have been found, these contain no numerals. The system described above is documented only in monumental contexts. The functions of the numerals included details of sacrifices or offerings to gods, quantities of booty obtained, numbers of military troops, and information on construction projects such as monuments and irrigation systems. The South Arabians did not use an enumerated dating system, nor do South Arabian coins contain numerical notation of any kind.

By the 2nd century BC, instances of the South Arabian numerals were normally preceded by the appropriate lexical numeral written out in full. While this aids modern scholars in their interpretation, doing so also removed any incentive to continue to use the system. By the 1st century BC, although the South Arabian scripts continued to be used, the numerical notation system had become extinct, and was not replaced until the 7th century AD, when the Islamic conquest brought alphabetic and later positional numerals to southern Arabia.

Berber

The Berbers live in North Africa and speak a set of closely related Afro-Asiatic languages. For most of their history, the Berbers have been a marginal people living on the periphery of larger polities (Carthage, Rome, and the Muslim nations), but have nonetheless retained considerable cultural independence. The Berbers developed a consonantal script on the model of that used in Punic Carthage, possibly as early as the 6th century BC, but definitely by the 3nd century BC (O'Connor 1996: 113). The classical Berber script was in continuous use until at least the 3rd century AD, and the Tifinigh script still used by the modern Tuareg for love letters, domestic ornamentation, and other purposes is clearly descended from it. There is no numerical notation system associated with either the classical Berber script or its modern descendant. Nonetheless, a distinct numerical notation system was used by traders in the Berber city of Ghadames (on the border of Algeria and Libya) in the 19th century.

This Berber system remains very poorly understood. It is not discussed in the general works of Menninger (1969), Guitel (1975), or Ifrah (1998), and is only described in a very small number of German papers (Rohlfs 1872; Vycichl 1952). Vycichl presents the system as described by two separate authors, Hanoteau and Si Mohammed Serif; I reproduce below both sets of numeral-signs in Table 4.17 (Vycichl 1952: 81-82).

Table 4.17: Berber numerals

	1	5	10	50	100	500	1000
Hanoteau	I	>	O	⊏	b		ɤ
Si Mohammed Serif	I	>	O	⊏	b	X̲	X̲̅

The system is purely cumulative-additive and written from right to left, with the decimal exponents repeated up to four times and the halved exponents only once in any numeral-phrase. Thus, 488 would be written as III>OOOCbbbb. Sometimes, groups of signs could be placed in two rows to save space, so that 44 could be written as IIOO
IIOO (Rohlfs 1872). In addition to these signs, Hanoteau claims that there was a fractional sign for 1/4, ▬, which could be stacked vertically to represent 1/4 (▬), 1/2 (▬▬), and 3/4 (▬▬▬) (Vycichl 1952: 81). The similarities between certain numerical signs and letters of the Berber consonantary (O=r, Ծ=f, X=s) are notable, but they do not correspond with the Berber lexical numerals in any obvious way. The two sets of numeral-signs are identical, except for the signs for 500 and 1000. It is possible that the two systems described above were actually both used, either in different contexts or in different times. However, it seems more likely that an error of interpretation created the discrepancy, especially because Hanoteau's 1000-sign is essentially identical to Serif's 500-sign. Which interpretation, if either, is the correct one, remains unknown.

The question of the Berber system's ancestor (if any) is still open. It is possible, but unlikely, that it was an entirely independent development, given the number of similarly structured systems in use in the Mediterranean. The Phoenician/Punic numerical notation system is quite different in its structure, given its lack of a sign for 5, its use of a special sign for 20, and its hybrid multiplicative-additive structure for expressing hundreds and thousands. Thus, while the Berber script is obviously based on a Punic model, its numerical notation system is not. The Berber system shares its structure with the systems of the Italic family. It is similar to the Lycian numerals; however, the 2500-year gap (and enormous geographical span) between the two regions makes such a hypothesis unlikely. The use of I for 1 and > for 5 is superficially similar to the Roman system. Moreover, Ghadames was an important trading post (Cydamus) under Imperial Roman control, and there are Roman numerals on some of the Latin inscriptions found there. However, the systems are written in different directions and have different signs for the higher values, and in any event, the chronological gap is too great to make this a plausible theory. Vycichl (1952: 83) suggests that the system owes its origin to diffusion from the South Arabian numerals. It has also been suggested that the Berber script may be somehow indebted to the South Arabian, mostly based on certain similar letters (O'Connor 1996: 112). If Hanoteau's list of signs is correct, the Berber system, like the South Arabian, lacks a sign for 500; furthermore, both systems use O for 10. However, the South Arabian system ceased to be used in the 1st century BC and was never used in Africa, so to accept this theory requires that we believe in a two thousand year unattested history for this system. The system having the most promise as an ancestor is the Arabico-Hispanic Roman variant (see above) used in a Spanish Inquisition document of 1576 (Labarta and Barceló 1988: 34). This system employed I, Ɔ, and O for 1, 5, and 10, was written from right to left, and was used in the same general region as the Berber system. Though three centuries is still a chronological gap that needs to be resolved, it is not nearly so great as the enormous leaps that need to be inferred to hypothesize alternate paths of diffusion.

The Berber system was used only for indicating the prices of trade goods. Ghadames has been an important trading post since Roman times, and remains so even today. Rohlfs (1872) learned about this system as a traveller in the Ghadames region, but only ascertained the meanings of the signs through great effort and negotiation. He thus indicated that the system's use was semi-cryptographic, and that it was employed to restrict the flow of information concerning prices to a limited group of Berber traders in order to give them an advantage over Arab traders. Yet the system is not especially difficult to decipher, and so I am unconvinced that this purpose was very important. I do not know of any surviving document that contains the Berber system (other than the reports of 19th century scholars), and there is no reason to believe that it continues to be used today.

Summary

The Italic numerical notation systems probably developed in the early 6th century BC with the invention of the Etruscan numerals, perhaps based on a previously existing system of tallies but possibly also influenced by the earlier Mycenean Linear B system. The heyday of this phylogeny was the 5th and 4th centuries BC, the era of classical Greek pre-eminence in the Mediterranean. However, of all the systems of this family, only the Roman numerals had any extensive use in the Christian era, as the Greeks had switched to the ciphered-additive alphabetic numerals (ch. 5) by the Hellenistic period. The remarkable persistence of the Roman system and the swift decline of other systems are not well explained by considerations of efficiency but rather by the changing political fortunes of their users. The use of Italic systems was for the most part limited to inscriptions and commercial marks, though the Roman numerals came to be used for an enormous variety of functions in different social contexts. While they were not used directly for arithmetic, there may be some connection between their cumulative-additive structure, their decimal bases, and quinary sub-bases and the abacus, which was used for computation in many parts of the Mediterranean and Europe.

All members of the Italic group of numerical notation systems share the following features: a) a cumulative-additive structure; b) a base of 10 with a sub-base of 5; c) the use of a single vertical stroke for the units. Some slight structural differences occasionally emerge at higher values, such as the use of implied multiplication in South Arabian and the hybrid multiplicative structure of later Roman numerals. Although the epichoric numerals of Argos, Nemea, and Epidaurus use dots rather than vertical strokes for units, they are clearly related to the Greek acrophonic numerals and must be considered part of this family.

The cultural history of some Italic systems is intermingled with those of the Hieroglyphic (ch. 2) and Levantine (ch. 3) families, making the construction of accurate cultural phylogenies more difficult. Because all three families are cumulative-additive, decimal, and used in the eastern Mediterranean, the affiliations of their systems can be difficult to discern. Often the three families can be distinguished on structural grounds: the Hieroglyphic systems all lack a quinary component, while the Levantine systems all have special signs for 20 and are multiplicative-additive above 100. This structural distinction can be confirmed independently by examining known patterns of historical contact. Despite the jumbled state of our present knowledge, each family is distinct, not only due to similarities in their systems' structure, but also as a result of the attested cultural connections among the societies in which they were developed.

Chapter 5: Alphabetic Systems

The families of systems I have discussed so far (the Hieroglyphic, Italic, and Levantine families) have mostly been cumulative-additive systems - those in which multiple signs are repeated within a single exponent of the base to indicate that those signs should be added. The next two chapters - the Alphabetic and South Asian systems - describe mainly ciphered-additive systems, which use, at most, a single sign for any exponent of the base to indicate the multiple of that exponent that is indicated: 1 through 9, 10 through 90, 100 through 900, and so on, in the case of a base-10 system. Specifically, most of the systems in this chapter are ciphered-additive. While ciphered-additive numeral-phrases are thus much shorter than cumulative-additive ones, ciphered-additive numerical notation systems require their users to be familiar with many more signs; where cumulative-additive systems normally have only one sign for each exponent of the base, decimal ciphered-additive systems require nine.

The family of systems that I will now discuss, which I call "alphabetic" numerical notation systems, in fact comprises a wide range of scripts, of which several are non-alphabetic. Many of the scripts whose numerals are discussed below, such as the Hebrew and early Arabic, are *abjads,* expressing primarily consonantal phonemes, and one, the Ethiopic Ge'ez script, is a syllabary, expressing consonant + vowel clusters. Leaving such details aside, the vast majority of systems in this family employ the letters of a script, *in a specified order*, to express the numerals using a ciphered-additive structure. I use the term "alphabetic" for this family being well aware that it does not apply to all the systems below.

This family is very extensive in both time and space. Systems of this family were known and used as far north as England, Germany, and Russia, as far west as Morocco, as far east as Iran, and as far south as Ethiopia. Its history spans over two thousand years, from the development of the Greek numerals around 600 BC to the present day. The major systems of this family are shown in Table 5.1.

Table 5.1: Alphabetic Numerical Notation Systems

	Greek	Coptic	Ethiopic	Gothic	Hebrew	Syriac	Arabic	Fez	Armenian	Georgian	Glagolitic	Cyrillic
1	Α ᾱ	ⲁ̅	፩	·𐌰·	א	ܐ	ا	[illegible]	Ա ա	ა	Ⰰ	А
2	Β β̄	ⲃ̅	፪	·𐌱·	ב	ܒ	ب	[illegible]	Բ բ	ბ	Ⰱ	В
3	Γ γ̄	ⲅ̅	፫	·𐌲·	ג	ܓ	ج	[illegible]	Գ գ	გ	Ⰲ	Г
4	Δ δ̄	ⲇ̅	፬	·𐌳·	ד	ܕ	د	[illegible]	Դ դ	დ	Ⰳ	Д
5	Ε ε̄	ⲉ̅	፭	·𐌴·	ה	ܗ	ه	[illegible]	Ե ե	ე	Ⰴ	Є
6	Ϛ ϛ̄	ⲋ̅	፮	·𐌵·	ו	ܘ	و	6	Զ զ	ვ	Ⰵ	Ѕ
7	Ζ ζ̄	ⲍ̅	፯	·𐌶·	ז	ܙ	ز	7	Է է	ზ	Ⰶ	З
8	Η η̄	ⲏ̅	፰	·𐌷·	ח	ܚ	ح	[illegible]	Ը ը	ჱ	Ⰷ	И
9	Θ θ̄	ⲑ̅	፱	·𐌸·	ט	ܛ	ط	[illegible]	Թ թ	თ	Ⰸ	Ѳ
10	Ι ῑ	ⲓ̅	፲	·𐌹·	י	ܝ	ي	[illegible]	Ժ ժ	ი	Ⰹ	І
20	Κ κ̄	ⲕ̅	፳	·𐌺·	כ	ܟ	ك	[illegible]	Ի ի	კ	Ⰻ	К
30	Λ λ̄	ⲗ̅	፴	·𐌻·	ל	ܠ	ل	[illegible]	Լ լ	ლ	Ⰼ	Л
40	Μ μ̄	ⲙ̅	፵	·𐌼·	מ	ܡ	م	[illegible]	Խ խ	მ	Ⰽ	М
50	Ν ν̄	ⲛ̅	፶	·𐌽·	נ	ܢ	ن	[illegible]	Ծ ծ	ნ	Ⰾ	Н
60	Ξ ξ̄	ⲝ̅	፷	·𐌾·	ס	ܣ	س	[illegible]	Կ կ	ჲ	Ⰿ	Ѯ
70	Ο ο̄	ⲟ̅	፸	·𐌿·	ע	ܥ	ع	[illegible]	Հ հ	ო	Ⱀ	О
80	Π π̄	ⲡ̅	፹	·𐍀·	פ	ܦ	ف	[illegible]	Ձ ձ	პ	Ⱁ	П
90	Ϙ ϙ̄	ϥ̅	፺	·𐍁·	צ	ܨ	ص	[illegible]	Ղ ղ	ჟ	Ⱂ	Ч
100	Ρ ρ̄	ⲣ̅	፻	·𐍂·	ק	ܩ	ق	[illegible]	Ճ ճ	რ	Ⱃ	Р
200	Σ σ̄	ⲥ̅	፪፻	·𐍃·	ר	ܪ	ر	[illegible]	Մ մ	ს	Ⱄ	С
300	Τ τ̄	ⲧ̅	፫፻	·𐍄·	ש	ܫ	ش	[illegible]	Յ յ	ტ	Ⱅ	Т
400	Υ ῡ	ⲩ̅	፬፻	·𐍅·	ת	ܬ	ت	[illegible]	Ն ն	უ	Ⱆ	Ѵ
500	Φ φ̄	ⲫ̅	፭፻	·𐍆·	תק	ܬܩ	ث	[illegible]	Շ շ	ფ	Ⱇ	Ф
600	Χ χ̄	ⲭ̅	፮፻	·𐍇·	תר	ܬܪ	خ	[illegible]	Ո ո	ქ	Ⱈ	Х
700	Ψ ψ̄	ⲯ̅	፯፻	·𐍈·	תש	ܬܫ	ذ	[illegible]	Չ չ	ღ	Ⱉ	Ѱ
800	Ω ω̄	ⲱ̅	፰፻	·𐍉·	תת	ܬܬ	ض	[illegible]	Պ պ	ყ	Ⱌ	Ѡ

	Greek	Coptic	Ethiopic	Gothic	Hebrew	Syriac	Arabic	Fez	Armenian	Georgian	Glagolitic	Cyrillic
900	Ͳ ϡ̄	ϣ̄	፱፻	·𐍊·	תתק	ܬܬܩ	ظ	[illegible]	Ջ ջ	შ	Ⱌ	ц
1000	͵ᾱ	ⲁ̲̄	፲፻		א׳	ܐ̗	غ		Ռ ռ	ჩ	Ⱍ	
2000	͵β̄	ⲃ̲̄	፳፻		ב׳	ܒ̗			Ս ս	ც	Ⱎ	
3000	͵γ̄	ⲅ̲̄	፴፻		ג׳	ܓ̗			Վ վ	ძ		
4000	͵δ̄	ⲇ̲̄	፵፻		ד׳	ܕ̗			Տ տ	წ		
5000	͵ε̄	ⲉ̲̄	፶፻		ה׳	ܗ̗			Ր ր	ჭ		
6000	͵ϛ̄	ⲋ̲̄	፷፻		ו׳	ܘ̗			Ց ց	ხ		
7000	͵ζ̄	ⲍ̲̄	፸፻		ז׳	ܙ̗			Ւ ւ	ჴ		
8000	͵η̄	ⲏ̲̄	፹፻		ח׳	ܚ̗			Փ փ	ჯ		
9000	͵θ̄	ⲑ̲̄	፺፻		ט׳	ܛ̗			Ք ք	ჰ		
10,000	$\overset{\alpha}{M}$		፼			ܐ̱				ჵ		
20,000	$\overset{\beta}{M}$		፪፼			ܒ̱						
30,000	$\overset{\gamma}{M}$		፫፼			ܓ̱						
40,000	$\overset{\delta}{M}$		፬፼			ܕ̱						
50,000	$\overset{\varepsilon}{M}$		፭፼			ܗ̱						
60,000	$\overset{\varpi}{M}$		፮፼			ܘ̱						
70,000	$\overset{\zeta}{M}$		፯፼			ܙ̱						
80,000	$\overset{\eta}{M}$		፰፼			ܚ̱						
90,000	$\overset{\theta}{M}$		፱፼			ܛ̱						
100,000	$\overset{\iota}{M}$		፲፼			ܝ̱						
1,000,000	$\overset{\pi}{M}$		፻፼			ܩ̱						

Historians of mathematics have thoroughly documented the history and functions of some of these systems, particularly the Greek, Hebrew, and Arabic ones, all of which were used extensively in mathematical contexts. In other cases, though, our knowledge of the histories of individual systems, and of the phylogenetic linkages among different systems, remains quite limited. In these cases, we will require much further research before we can reach any definite conclusions. It is part of my goal in this chapter to illuminate areas of study where our knowledge is less than perfect to draw attention to the need for further specialized research on such topics. In particular, it should be obvious from the above table that, while these systems are ciphered-additive for the most part, they are certainly not structurally identical. From these structural differences, rather than the paleographic curiosities of the signs of various systems, we can learn the most about this family.

Greek alphabetic

In chapter 4, I discussed the cumulative-additive Greek acrophonic numerals, which were given their name because the letters used are the first letters of the appropriate Greek numeral words. This system is entirely independent of the other, and more frequently encountered system, which is sometimes called the "Ionic" or "Milesian" system due to its origin in western Asia Minor, but which I simply call the alphabetic system, the term most commonly used today. The Greek alphabetic numerals are one of the most frequently discussed numerical notation systems, being of interest to both epigraphers and historians of mathematics, yet they remain fundamentally misunderstood.

While the Greek alphabet was developed after a Phoenician model, probably in the 9th or 8th century BC, none of the earliest Greek inscriptions contains numerical notation; thus, the debates on the time of the origin of the alphabet can safely be ignored when examining the numerals (cf. McCarter 1975; Swiggers 1996 for a review of the evidence). The first examples of the alphabetic numerals date to the 6th century BC and are written using the letters of the archaic Greek script used in Ionia and the Ionian cities of Caria, such as Miletus, as shown in Table 5.2.

Table 5.2: Greek alphabetic numerals (archaic)

	1	2	3	4	5	6	7	8	9
1s	Α	Β	Γ	Δ	Ε	Ϝ	Ι	Η	⊕
10s	Ι	Κ	Λ	Μ	Ν	Ξ	Ο	Π	Ϙ

100s	P	ξ	T	V	Φ	X	Ψ	Ω	M

The system was purely ciphered-additive and decimal, and was usually written from left to right, though right-to-left and boustrophedon inscriptions are not unknown. Thus, 562 would be written as ΦΞB. The numeral-signs are archaic variants of the 24 familiar letters of the Greek alphabet, plus three special signs called *episemons*: vau or digamma (6), qoppa (90), and san or sampi (900), which were added to reach a full complement of 27 signs for all the values from 1 through 900, enabling any natural number less than 1000 to be written. Vau and qoppa were occasionally used phonetically in the Ionic script, with the rough values of [v] and [k], while san appears to have been borrowed from Phoenician *sade* [ts], though it may have occasionally been used in archaic Greek with a similar phonetic value (Swiggers 1996: 265-66). By the time the alphabetic numerals were developed, san had lost the place between pi (Π) and qoppa (Ϙ) that it had held in the Phoenician script, and was placed at the end of the system, with the value of 900. There are very few examples of alphabetic numerals from this early period, and all of them express values under 1000, so we have no way to establish whether higher values could be represented.

In classical and Roman Greece, the familiar Greek alphabet developed out of the archaic regional (or "epichoric") variants and the alphabetic numerals developed along with them, retaining their order and numerical values but coming to assume their modern (majuscule) forms. In addition, starting in the middle of the 5th century BC, we have evidence of two new techniques used to express higher values. For multiples of 1000, a small slanting mark (known as a *hasta*) was placed to the left and below a sign for 1 to 9 to indicate its value should be multiplied by 1000; thus, Γ means 3 but ͵Γ means 3000 (Threatte 1980: 115). Values above 10,000 are rarely encountered except in mathematical works, and individual mathematicians used different methods to do so. The most common method, used by Aristarchus, involved placing a small alphabetic numeral-phrase (less than 10,000) above a large M character (=*myriades*) to indicate multiplication by 10,000 (Heath 1921: 39-41).[1] Thus, 3,000,000 would be expressed with only two signs, as $\overset{\psi}{\mathrm{M}}$. This allowed any number less than 100 million to be easily expressed. In papyri of the Roman period, a large number of variant multiplicative signs for 1000 and 10,000 were used, most notably ∩ for 10,000, which was the normal form starting in the 2nd century AD (Brashear 1985). The entire system thus came to appear as shown in Table 5.3, using the classical letters used at Athens.

Table 5.3: Greek alphabetic numerals (classical)

	1	2	3	4	5	6	7	8	9
1s	Α	Β	Γ	Δ	Ε	Ϝ	Ζ	Η	Θ
10s	Ι	Κ	Λ	Μ	Ν	Ξ	Ο	Π	Ϙ
100s	Ρ	Σ	Τ	Υ	Χ	Φ	Ψ	Ω	Ϡ
1000s	͵Α	͵Β	͵Γ	͵Δ	͵Ε	͵Ϝ	͵Ζ	͵Η	͵Θ
10,000s	$\overset{\mathrm{A}}{\mathrm{M}}$	$\overset{\mathrm{B}}{\mathrm{M}}$	$\overset{\Gamma}{\mathrm{M}}$	$\overset{\Delta}{\mathrm{M}}$	$\overset{\mathrm{E}}{\mathrm{M}}$	$\overset{\text{Ϝ}}{\mathrm{M}}$	$\overset{\mathrm{Z}}{\mathrm{M}}$	$\overset{\mathrm{H}}{\mathrm{M}}$	$\overset{\Theta}{\mathrm{M}}$

This system, as it came to be used in classical Greece, is thus ciphered-additive for values under 1000, and thereafter is multiplicative-additive at two different levels: firstly, through the use of a hasta to indicate multiplication by 1000, and then through the use of an M to indicate multiplication by 10,000. It might be asked why the Greeks would not simply continue the series using 10-90 and 100-900 preceded by the hasta (͵Ι = 10,000; ͵Κ = 20,000; ͵Λ = 30,000, etc.). It is possible that these two separate levels represent progressive steps in the system's development, with the second (myriads) series being a later development. I think it more likely that this feature is a clue to the alphabetic numerals' history. If, as I contend, the Greek numerals were derived from the demotic

---

[1] Heath also discusses techniques such as that of Heron's *Geometrica*, where two dots placed over a sign indicate multiplication by 10,000, that of Apollonius, using "tetrads", turning the system into a mixed base-10/10,000 ciphered-additive system, and that of Nicholas Rhabdas, a 14th century scholar who used Heron's technique, except that additional pairs of dots above a number indicated successive exponents of 10,000. None of these systems was ever widely used.

numerals, then it would be reasonable for the Greeks to adopt the multiplicative principle at the same level as in the demotic numerals, namely 10,000. However, because the Greek alphabet only has 24 letters, and requires three episemons to reach the 27 signs needed to get as high as 900, it would not have been feasible to find nine extra signs for the values 1000-9000. Consequently, the inventor(s) of the alphabetic numerals may have had the idea of using multiplication for the thousands values as well as the ten thousands. The only remaining problem is to explain why the Greeks, recognizing this irregularity, did not simply start using multiplication at the thousands level and abandon the higher multiplicative series.

The alphabetic numerals were generally written in descending order, with the highest values on the left. In many cases, however, numbers between 11 and 19 were written with the 10-sign (Ι) following the unit-sign, to correspond with the way that the ancient Greek lexical numerals were formed: *hendeka, dodeka, treis kai deka, tettares kai deka,* and so on. For instance, Threatte (1980: 114) provides a number of examples from Attica where ΓΙ, ΖΙ, ΗΙ, and ΘΙ appear for 13, 17, 18, and 19 even though those numeral-phrases would normally be written in the reverse order. In the Roman period or later, the order of signs became more rigidly fixed in highest-to-lowest order.

Classical Greek alphabetic numerals were sometimes distinguished from the rest of the text with special signs, most commonly a horizontal stroke above the numeral-phrase, but occasionally with dots placed to either side of it. Because the numerals could easily be confused with written words, this delineation served to distinguish numerals within a block of text. One of the problems in identifying earlier Greek alphabetic numerals is the lack of such marks, meaning that any single letter could be an alphabetic numeral or a non-numerical label. Regardless, even in later periods, specially denoting numerals was not a universal practice, and numerals frequently appear without any indicator mark whatsoever (Threatte 1980: 115).

In most Greek monumental inscriptions, no fractional values are found save for monetary units, for which separate signs existed for different units of currency and fractions thereof (Tod 1950: 134). However, these are not part of the standard alphabetic numerals, but rather are associated with the acrophonic numerals in origin and means of expression. In mathematical and literary texts, an entirely different system was used in which two small accents or strokes placed to the right of a numeral indicated a given unit-fraction (Thomas 1962: 43). Special signs existed for 1/2 (∠' and C') and 2/3 (ω'), in addition to standard unit-fractions (Thomas 1962: 45). This system is almost certainly akin to the unit-fraction systems of the Egyptian hieratic and demotic numerals, which also used special signs for 1/2 and 2/3. From the 2nd century AD onward, the requirement of using only unit-fractions was lifted, and fractions were expressed with both numerators and denominators using alphabetic numerals. Later in this chapter, I will discuss the Greek astronomical fractions, which combine the alphabetic numerals with sexagesimal structures borrowed from the Babylonians.

The debate regarding the origin of the alphabetic numerals has not progressed in a century, and, as Johnston indicates, the study of the early history of the Greek numerals (both alphabetic and acrophonic) has generally been ignored in favour of limited studies of regional variations that developed much later (1979: 27). When they have considered the topic, classical epigraphers have assumed that the alphabetic numerals were independently invented, without considering the possibility that the system has an external origin – an error that neglects an extremely likely ancestor. Before addressing this issue, however, I must begin by reviewing the early history of the numerals, as well as some older theories of their origin that have now been discounted.

In the fifth century AD, the Neoplatonist philosopher Proclus suggested that the Greek alphabetic numerals were modelled on an earlier Phoenician system, since the Greek alphabet was borrowed from a Phoenician ancestor and because many Semitic scripts used consonantal signs as numerals (Brunschwig and Lloyd 2000: 388). This theory was widely accepted as late as the mid-nineteenth century. Yet it is now firmly refuted, as no Semitic consonantal or alphabetic numerical notation systems existed before the 2^{nd} century BC, and all such systems were based on a Greek model rather than the other way around (Gow 1883). The idea that the alphabetic numerals must have been developed in the late 8^{th} century BC, shortly after the development of the alphabet, has now also been rejected due to a lack of material evidence, though it enjoyed some popularity due its espousal in Larfeld's *Handbuch der griechische Epigraphik* (Larfeld 1902-1907). Similarly, the once-popular theory of a very late origin (late 4^{th} or even 3^{rd} century BC) cannot be sustained in light of evidence from much earlier periods (Gow 1883).

The first epigraphic evidence for alphabetic numerals comes from a vase, dating to around 575 BC, found at Corinth, which contains the inscription "SYM I", which Johnston reads as "mixed batch of 7" (Johnston 1973: 186). There is good evidence from Attica and Corinth for the system's use on mercantile vases, especially in the late 6^{th} and early 5^{th} centuries BC (Hackl 1909). Yet it is unlikely that the numerals actually developed in either of these localities. Rather, the numerals probably developed in western Asia Minor, in the regions of Ionia and Caria, especially the cities of Miletus[2] and Halicarnassus, where several 6^{th} and 5^{th} century BC instances of the numerals have been found (Heath 1921: 32-33). All the early examples of the alphabetic numerals, even those found outside Asia Minor, are written with the Ionic script, which was used in Ionia, Caria, and various Ionic colonies throughout the Mediterranean. The predominance of alphabetic numerals of Ionic scripts reflects the predominance of Ionia in regional and international commerce during the 6^{th} and early 5^{th} centuries BC.

That the script is Ionic or Milesian in origin is confirmed by two facts. Firstly, the alphabet used in Ionia and at Miletus had characters for two of the episemons, vau and qoppa, that were not both present in any other Greek alphabet at the same time as other

---

[2] Miletus, from whence the adjective "Milesian", was the most important Ionian city in Caria, the region of Asia Minor immediately to the south of Ionia proper.

characters (phi, chi, psi, and omega) that were an integral part of the system (Jeffery 1990: 327). Thus, no other alphabet is likely to have developed numerals in the order universally found with the alphabetic numerals (vau = 6, qoppa = 90, chi=500, phi=600, psi=700, omega=800). The third episemon, san, was not a regular part of the Milesian or any other Greek alphabet at the time, and thus was placed at the end of the series.

Secondly, and perhaps more remarkably, after a period of Ionian cultural dominance between 575 and 475 BC, when alphabetic numerals were commonly found, alphabetic numerals are found only rarely in a period starting in 475 BC and lasting around 150 years (Johnston 1979: 27). During this period, the height of Greek achievement, Athens came to the forefront as an Aegean power, and the acrophonic numerals used in Athens were used in most Greek-speaking areas, while Ionia's power waned after the Milesian-led Ionian revolt against Achaemenid Persia of 499 to 494 BC. The system did not disappear entirely; at Halicarnassus, there is solid evidence of the continued use of the system between 450 and 350 BC (Heath 1921: 32-33). In addition, a very curious inscription from Athens (IG I^2 760) from the middle of the 5th century BC contains a long series of alphabetic numerals (written with Ionic letters). That the disappearance of the numerals corresponds with the decline of Ionia is further evidence that the system originated in Asia Minor rather than in Greece proper.

Most classicists accept that the alphabetic numerals were an early 6th century BC invention in western Asia Minor (Johnston 1979; Jeffery 1990). The question that remains unasked, however, is whether this development was stimulated, directly or indirectly, by some other system in existence at the time. I believe that the structure of the Greek alphabetic numerals was borrowed directly from the Egyptian demotic numerals, only using alphabetic signs as numeral-signs. Neither Egyptologists nor classicists have examined this theory; only Boyer (1944: 159) has seen the similarity between the two systems as indicative of a historical connection, and his paper was not primarily oriented towards such an argument.

One of the difficulties in tracing the origin of the alphabetic numerals is that, because they are the first to use phonetic signs as numeral-signs, it is impossible to use paleographic evidence of their similarity to any earlier system – there simply will be no such similarity, because there are no earlier alphabetic systems. Yet, in almost all major structural aspects, the systems are identical. They are both ciphered-additive, base-10 systems. While it has yet to be established whether the alphabetic numerals used multiplicative notation at an early date, it is suggestive that both systems are multiplicative-additive above 10,000. The alphabetic system is also multiplicative for the thousands, which is not the case for the demotic numerals. However, as mentioned earlier, the most obvious step after the use of the "hasta + units" multiplicative formation for the thousands would be to use "hasta + tens" and "hasta + hundreds" for the ten thousands and hundred thousands, respectively. There must be a reason why the Greeks began a new multiplicative series using 10,000 as the multiplicand. One possibility is that the demotic numerals, which were multiplicative only for 10,000, provided a model for doing so.

Furthermore, Greek arithmetical techniques for dealing with fractions show a remarkable similarity with the Egyptian unit-fraction (1/x) tradition of computation (Knorr 1982; Fowler 1999). Both the demotic numerals and the early Greek alphabetic numerals used unit fractions formed by placing a small mark above a given integer to indicate the appropriate unit fraction. Furthermore, both used alternative non-unit fractions for specific fractional values, although this is more prevalent in the demotic numerals than in the Greek alphabetic numerals, which only did so for 1/2 and 2/3. Historians of mathematics are unanimous that the Greeks borrowed the unit-fraction technique from the Egyptians, and I see no reason to doubt that the Greek use of special signs for 1/2 and 2/3 is also a result of Egyptian influence.

The historical connections between Egypt and Greece are even more convincing than the structural similarities between the systems. The demotic numerals were the predominant ones in use in Egypt (especially Lower Egypt) in the early 6th century BC. This was just the time when Greeks were starting to encounter Egyptians in large numbers for the purposes of international trade. Most notable among the Greek traders in Egypt were colonists from Miletus, who had set up an important *emporion* (port of trade) at Naukratis in the western Nile delta in the 7th century BC. Naukratis quickly became the central locus for trade and cultural contact between Greece and Egypt, a position that it held until the Ptolemaic era. Inscriptions in the Ionic Greek script have been found at Naukratis dating as early as 650 BC (Heath 1921: 33). It should be noted, however, that no known inscriptions from Naukratis contain alphabetic numerals, and there are later (4th century BC) inscriptions with acrophonic numerals (Gardner 1888). That there was enormous trade going on between the Aegean and Egypt during this period can hardly be disputed. Since the earliest examples of the alphabetic numerals are from vases and jars used to hold commercial goods, the context of the system's development was probably in mercantile activity.

Thus, the preponderance of evidence suggests that the Greek alphabetic numerals are descended from the demotic numerals of the early 6th century BC. As a more remote possibility, the late hieratic numerals are a possible ancestor, though by the Late period, the hieratic numerals were mostly used in Upper Egypt, where there were few Greeks, and tended to use cumulative-additive rather than ciphered-additive numeral-phrases. The alternative to the hypothesis of Egyptian borrowing is that the Ionians independently developed a ciphered-additive, decimal numerical notation system within a few decades of coming into contact with Egyptians in large numbers, founding a colony at Naukratis, and no doubt being exposed to the demotic numerals that were widely used for administration and commerce throughout Lower Egypt. While the lack of paleographic evidence from the numeral-signs makes it difficult to prove the case, the presumption that the Greek alphabetic numerals were independently invented ought to be replaced by a working hypothesis of direct diffusion from the demotic numerals.

This should not be taken as a denial of the Greeks' inventiveness, because the alphabetic numerals have several distinctive properties. Firstly, while some of the demotic numeral-signs use the cumulative principle, the alphabetic numerals use purely ciphered signs. Secondly, as mentioned above, the alphabetic numerals use the multiplicative principle for 1000-9000, obviating the need for nine more signs for those values, as one could simply write a hasta before a unit-sign. Finally, by virtue of the fact that most of its numeral-signs would already be understood (and their order known) by literate Greeks, the alphabetic numerals, in contrast to the demotic, did not require the learning of an enormous new set of signs. Rather, only the numerical values attached to the signs needed to be learned, and anyone who already knew the order of the alphabet could determine the signs' values as long as the episemons were taken into account. The often-mentioned "weakness" of the alphabetic numerals, that too many signs needed to be learned, is largely illusory, even when comparing the system to our own. In learning to read and write, Western pupils must learn 26 alphabetic signs (in their proper order) *plus* 10 digits in order, making 36 total signs in two separate series, while the ancient Greeks needed only to learn 27 alphabetic signs and two auxiliary signs (͵ and M), and thus only needed 29 total signs in one series. Diffusion from Egypt does not imply that there is nothing special or interesting about these local Greek developments.

The resurgence of the alphabetic numerals in Greece around 325 BC corresponds almost exactly with the rise of the Ptolemies in Egypt. In this renewed period, some of the earliest instances of the numerals come from Egypt. The Hibeh papyrus, a Greco-Egyptian astronomical document dating to around 300 BC, is one of these early instances (Grenfell and Hunt 1906; Fowler and Turner 1983). Similarly, coins dating to 266 BC indicating the regnal year of Ptolemy II Soter are, to my knowledge, the first coins bearing any ciphered-additive numerals (Tod 1950: 138). It is interesting that while these examples come from Egypt, we have no record of the alphabetic numerals' use in Egypt during the interlude of the 5th and 4th centuries BC. The evidence, at present, is simply too scanty to conclude what specific stimulus caused the rejuvenation of the alphabetic numerals.

From the 3rd century BC onwards, the alphabetic numerals began to be preferred over the acrophonic numerals throughout most of the Greek-speaking world, with only Athens retaining the acrophonic system until around 50 BC (Threatte 1980: 117). While the acrophonic numerals used in different city-states varied quite widely, the alphabetic numerals had no regional variants. As such, they could easily be used as an effective instrument of cross-cultural communication and trade among diverse regions of Greece (Dow 1952: 23). Whereas Greece before Alexander was highly fragmented, rendering the development of a universal Greek numerical notation system unlikely, Alexandrine and especially Roman Greece provided a suitable environment for the development of a single pan-Hellenic notation. That the system was a very concise way to represent numbers, and that it relied on alphabetic symbols that were themselves invariant throughout Greece by this period in history, cannot have hurt this process.

In the 6th and 5th centuries BC, the alphabetic numerals were no more than a system for labelling mercantile containers. All the early instances of the system's use are from marked vases and potsherds. Even then, most numerals on vases are acrophonic or other cumulative-additive Greek numerals, not alphabetic ones. In these very early contexts, the alphabetic numerals, like the acrophonic ones, were used for cardinal quantities, particularly of money, weights and measures, and discrete quantities of commodities, the sorts of numerical expressions likely to be found in inventories and decrees. From the 3rd century BC onward, though, when the alphabetic system became the predominant one throughout the Greek world, the numerals were used in a much wider range of contexts. In contrast to the acrophonic numerals, which are found solely on ceramic vessels and stone, alphabetic numerals are found, in addition, in manuscripts of various sorts as well as on coins. As described by Tod (1950: 130-134) and Threatte (1980: 115-116), the functions for the alphabetic numerals include:

a) cardinal quantities of commodities, persons;
b) phrases indicating lengths of time in days, months and years;
c) monetary values (denarii, drachmas, and obols);
d) weights, measures, and distances;
e) ordinal numerical adjectives and adverbs;
f) ordinal dates, e.g. to indicate a specific year in the tenure of an archon.

I have not yet discussed the use of the Greek alphabetic numerals for mathematics. During the early history of the numerals (6th to 4th centuries BC), we have no evidence that the alphabetic numerals were used for mathematics, and there is evidence that early writers used the acrophonic numerals along with a pebble-board or abacus (Lang 1957). This situation changed once the alphabetic numerals began to be used more widely. As compared to Egypt, where we have only a handful of surviving mathematical texts, many surviving Greek manuscripts from 300 BC until 1450 AD use the alphabetic numerals. At the beginning of this period, we find texts such as the Greco-Egyptian Hibeh papyrus, which I will discuss below. The alphabetic numerals were used by all Greek mathematicians, beginning with Archimedes and Apollinius, whose use of the numerals in their 3rd century BC mathematical works prompted later scholars to follow suit. One would think, as Boyer comments, that the adoption of the alphabetic numerals by two such prominent mathematicians would curb the criticism of modern scholars as to the system's usefulness for mathematics (Boyer 1944: 160). Nevertheless, the subject of the inferiority of the alphabetic numerals, not only to ciphered-positional systems such as our own but also to cumulative-additive systems such as the acrophonic numerals, has been a popular topic in the history of numeration (cf. Boyer 1944: 160-166). The system's brevity of expression may be counteracted somewhat by the large number of signs needed to learn it and the fact that, for instance, there is no resemblance among the signs for 5, 50, and 500. Yet the interminable arguments over the pros and cons ignore two important facts. Firstly, the numerals were used and promoted by many Greek mathematicians, producing insights that would not be equalled for centuries. The only "evidence" for their insufficiency for mathematical purposes is the assertion of some modern historians of mathematics that they appear cumbersome. In the only instance of which I am aware of a modern scholar actually attempting to learn and use the numerals - Paul Tannery's study in the 1880s - the system fared very well. Tannery found that calculation with alphabetic numerals took little more effort than with Western numerals, with which Tannery no doubt, despite all his efforts, would have had far greater experience and familiarity (Boyer 1944: 160-161). Secondly, although many mathematical texts were written with the alphabetic numerals, many more texts containing the numerals served non-mathematical functions. While historians of mathematics are naturally going to be interested in the numerals' use in mathematical contexts, we should not fall into the trap of thinking that this limited function tells us much about their usefulness overall.

Throughout their history, the Greek alphabetic numerals were used primarily in Greek-speaking areas or in regions under the control of Greek speakers. During its early history, the system was geographically restricted to the eastern Mediterranean, particularly the Aegean. Its widespread use in Ptolemaic Egypt and Seleucid Persia marked its maximal geographic spread. After the Roman conquest of Egypt and the rise of the Parthian Empire, Greek numerals were found in Greece, Asia Minor, and the Levant, though the Romans never used them for administrative purposes. The use of alphabetic numerals for administration recommenced upon the division of the Roman Empire into west and east. From the 4th century AD onward, they were used as the primary numerals of administration, law, literature, and mathematics in the Eastern Roman Empire. Whenever and wherever the Greek alphabet was used in the Middle Ages, the alphabetic numerals followed. Additionally, the Greek alphabetic numerals were the most common system used in Arabic papyri for several centuries after the Islamic conquest for recording the results of financial transactions, even after the invention of the Arabic abjad numerals (Grohmann 1952: 89).

In general, however, when the opportunity arose for cultures in contact with Greece to adopt numerical notation systems, they usually substituted the letters of their own scripts for the Greek signs. The first people to do so were the Israelites, in the late 2nd century BC, when they developed the system still used with the Hebrew script today. Thereafter, the Greek alphabetic numerals did not spread widely until near the end of Roman imperial domination and the beginnings of Byzantine Greek culture. From 350 to 500 AD, however, the system diffused widely as the Eastern Roman Empire began to exert its influence to the north, south, and east. In the mid-4th century, the Goths adopted alphabetic numerals along with their Greek-influenced script. In regions of Africa under Greek influence, the Coptic script of Egypt and the Ge'ez script used of Ethiopia both developed alphabetic numerical notation systems based on a Greek model around the same time. In Armenia and Georgia, right on the border of the Eastern Roman Empire, scripts and accompanying alphabetic numerals developed in the 5th century AD at around the time they were Christianized. Shortly thereafter, perhaps about 500 AD, the Syriac script was undergoing many changes, including a shift from an earlier cumulative-additive numerical notation system (ch. 3) to one based on the Greek. It is also probable that the Arabic abjad numerals used following the Islamic conquest of the Middle East were at least partly derived from the Greek alphabetic system. The final direct descendants of the alphabetic numerals were the Glagolitic and Cyrillic numerals, which developed in the late 9th century AD, under the auspices of the missionary work and script development of Cyril and Methodius in Slavic regions.

It is possible that the Greek numerals are ancestral to the Brāhmī numerals, which were used from the late 4th century BC onwards in India, and which themselves eventually gave rise to Western numerals. The Brāhmī numerical notation system is ciphered-additive and decimal, and used a variety of the multiplicative principle. The chronology of its invention, corresponding almost exactly with the Alexandrine conquests and journeys in India, is also suggestive. On the other hand, the Brāhmī system is more similar in structure and numeral-signs to the Egyptian demotic numerals than to the Greek alphabetic numerals. I will discuss the origins of Brāhmī numerals more thoroughly in Chapter 6.

The Greek alphabetic numerals directly gave rise to more descendants than almost any other system in history. This is in part a factor of their longevity, but other systems, such as the Egyptian hieroglyphs, were used over a much longer duration, yet generated few direct descendants. Other long-lived systems, such as the Roman numerals, spread very widely over large parts of the world due to Roman imperial power, but they were often accepted by colonized or subordinate societies unchanged, and did not replace indigenous systems entirely. Because the Greek system was alphabetic, cultures borrowing the principle of alphabetic numerals tended to modify the signs to fit their own scripts (whether alphabets or consonantaries) rather than adopting the Greek alphabetic numerals directly, and also made minor structural changes to the system.

The eventual fate of the Greek numerals was directly tied to the fortunes of the Byzantine Empire, the only major polity in which the numerals were used throughout the Middle Ages. In the early Middle Ages, when the Empire's fortunes were prosperous, the numerals were widely used throughout Greece, the Balkans, Egypt, the Levant, and Asia Minor, and were incorporated into the learning of all European mathematicians. For instance, they were known to the English scholar Bede, who described them in his *De temporum ratione* (The Reckoning of Time) in the early 8th century AD (Wallis 1999). However, by 1300, the geographical extent of the numerals' use was more limited than it had been since the Ptolemies, and mathematicians were already using something like our modern Western numerals under the influence of Arab learning in Spain and Italy. In the Byzantine Empire, mathematicians used Arabic positional numerals in marginal notes on Euclid's *Elements* in the 12th century (Wilson 1981), although the first major Byzantine mathematician to recommend the switch to the Arabic numerals was Maximus Planudes

(c. 1260-1310). In the great debate between the "abacists" favouring Roman numerals and abaci and the "algorithmicists" favouring the use of Western numerals, the Greek alphabetic numerals, used by all the great mathematical minds of antiquity, did not rate a mention. In a few 15th century mathematical texts, the Greek alphabetic numerals were transformed into a ciphered-positional system along the model of the Arabic or Western numerals, using only the first nine letters to indicate the units, and adding a dot to indicate a zero position (Menninger 1969: 273-4). In 1453, with the fall of Constantinople, the Greek numerals ceased to be used administratively. Nevertheless, they continued to be used thereafter for restricted purposes, such as paginating of religious and scholarly texts and enumerating ordinal lists, just as the Roman numerals were used in Western Europe. This limited use of the alphabetic numerals continues today, even though they have not been in regular use for over five centuries.

Coptic

The Coptic alphabet originated in Egypt in the 4th century AD and was largely based on the Greek alphabet, but used six additional characters taken from the demotic script to express the phonemes of the Egyptian language, which it was designed to represent. Unlike any of the earlier Egyptian scripts, Coptic is written from left to right and has signs for vowels. The adoption of the Coptic script was accompanied by the introduction of a ciphered-additive numerical notation system based on the model of the Greek alphabetic numerals (Megally 1991; Messiha 1994). In the system's classical form, its numeral-signs were as shown in Table 5.4.

Table 5.4: Coptic numerals (classical)

	1	2	3	4	5	6	7	8	9
1s	ⲁ̅	ⲃ̅	ⲅ̅	ⲇ̅	ⲉ̅	ⲋ̅	ⲍ̅	ⲏ̅	ⲑ̅
10s	ⲓ̅	ⲕ̅	ⲗ̅	ⲙ̅	ⲛ̅	ⲝ̅	ⲟ̅	ⲡ̅	ϥ̅

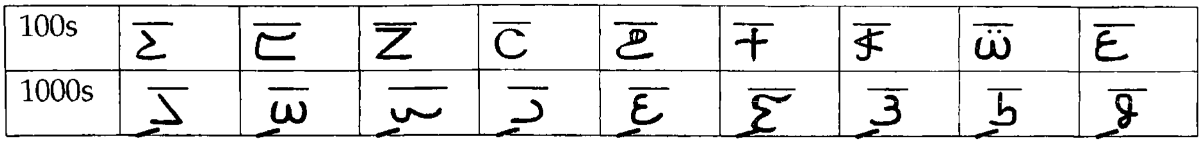
100s	ⲣ̄	ⲥ̄	ⲧ̄	ⲩ̄	ⲫ̄	ⲭ̄	ⲯ̄	ⲱ̄	ϣ̄
1000s	ⲁ̄	ⲃ̄	ⲅ̄	ⲇ̄	ⲉ̄	ⲋ̄	ⲍ̄	ⲏ̄	ⲑ̄

The numerals, like the script itself, were written from left to right. The system is ciphered-additive and decimal, and so 6085 would be written ⲋ̄ⲡ̄ⲉ̄. The numeral-signs are clearly derived from the Greek uncial signs used between the 4th and 9th centuries AD. In addition, like the Greek alphabetic numerals, a horizontal stroke above the numeral-phrase indicates that it is a numeral rather than a word, and a slanted subscript stroke under a unit-sign (the Greek *hasta*) indicates multiplication by 1000 (Megally 1991: 1821). There is no known sign or multiplier for 10,000 or higher values for these numerals.

The classical age of Coptic lasted from the 4th to the 10th centuries AD, during which time the script and numerals were used extensively, surviving the 7th century AD Muslim conquest of Egypt. There may have been a geographical division in the frequency of their use, with northern Egyptian scribes using them frequently, while southern writers tended to write out numbers using lexical numerals (Till 1961: 80). While the Coptic numerals were generally used in formal manuscripts, the ordinary cursive Greek alphabetic numerals were used for calculation and administration, possibly because the Coptic numerals, being uncials without tails, were less practical for rapid writing (Megally 1991: 1821). Furthermore, in the 10th century AD, when the Coptic script was being replaced by Arabic for most administrative purposes, a unique Coptic cursive numerical notation system developed, known as "numerals of the Epakt" (Messiha 1994: 26). This system is shown in Table 5.5.

Table 5.5: Coptic 'numerals of the Epakt'

	1	2	3	4	5	6	7	8	9
1s	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10s	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]



This system is structurally similar to the classical system, but the numeral-signs are cursive minuscule letters rather than uncial majuscule ones. Many of these signs bear little or no resemblance to the classical signs. Some of them are probably taken from the signs of the Arabic abjad numerals, which I will describe below, while others are of indigenous development. This system also uses two stages of multiplication (at 1000 and 10,000), similar to the Greek alphabetic system. A horizontal stroke and two dots placed below a sign indicated multiplication by 10,000 (Sesiano 1989: 64). A 15th century multiplication table (now in Istanbul) includes instructions on writing 'numerals of the Epakt', and indicates that this sign for 10,000 could be used in conjunction with any of the 27 letters, thus allowing any number less than ten million to be expressed (Sesiano 1989: 54-55). Given that the Greek numerals would have been well known in Egypt even at a relatively late date, it is impossible to establish when exactly the 'numerals of the Epakt' arose. It may be that the two-stage multiplicative principle at 1000 and 10,000 existed even in the earlier uncial numerals, and that we simply have no paleographic evidence to confirm this.

The paleographic relation between the first and second Coptic systems remains unclear, although some of the signs are clearly related. It appears that the Epakt signs were developed in the 10th century to aid in Arab administration in Egypt and continued to be used as late as the 17th century (Messiha 1994: 26). These numerals were often used in bilingual Arabic-Coptic documents, suggesting that the Arabs were making concessions to local administrators. This situation is quite extraordinary, given that some Egyptian Arabs by this time were employing the ciphered-positional Arabic numerals used today. This suggests that the advantages of ciphered-positional systems over ciphered-additive ones, such as the Coptic numerals, may not have been evident or important at the time.

It is unclear whether the Ethiopic numerals (used to write the Ge'ez language from the 4th century AD onward) were based directly on the Greek alphabetic numerals or derived through a Coptic intermediary. While the Ethiopic system is generally said to derive directly from Greek, the Coptic uncial letter-signs are similar enough to the Greek to render such a determination premature. Otherwise, the only system occasionally thought to be descended from Coptic is the so-called "Fez numerals" used in North Africa (see below). Again, however, this determination is premature, as either the Greek alphabetic numerals or the Arabic abjad are possible ancestors for the Fez numerals.

While the primary function of Coptic numerals has always been religious, given the script's use in the Coptic Church, their administrative and arithmetical functions should not be discounted. Despite the control of the population of Egypt by a succession of foreign powers, the use of Coptic for dating documents, accounting, commerce, and arithmetic continued as late as the 14th century, while the cursive Epakt numerals ceased to be used only in the 17th century. Furthermore, the classical uncial numerals are still used today in Coptic Christian liturgical texts for pagination and stichometry. Yet, for most ordinary purposes, either Western or Arabic numerals are preferred by those familiar with Coptic.

Ethiopic

The Ethiopic script developed in the 4th century AD, primarily on the model of the Minaeo-Sabaean script used in South Arabia, but also influenced by the Greek and Coptic alphabets used to the north. It was used (and continues to be used) for writing various languages of Ethiopia, especially the Ge'ez liturgical language of the Ethiopian church and modern Amharic. The script is an alphasyllabary, in which each individual sign represents a consonant + vowel cluster and in which the direction of writing is always left to right.

From its earliest appearance around 350 AD, on inscriptions from the kingdom of Aksum, the numerical notation accompanying this script is not the cumulative-additive one used in the South Arabian inscriptions, but a hybrid ciphered-additive and multiplicative-additive system based on the Greek alphabetic numerals. The numerical notation system used in the Aksum inscriptions (on which the signs for 70 and 90 are not found) is shown in Table 5.6 (Ifrah 1998: 247).

Table 5.6: Ethiopic numerals (Aksumite)

	1	2	3	4	5	6	7	8	9
1s	፩	፪	፫	፬	፭	፮	፯	፰	፱
10s	፲	፳	፴	፵	፶	፷		፹	
100s	፻	፪፻	፫፻	፬፻	፭፻	፮፻	፯፻	፰፻	፱፻

The modern Ethiopic script uses a system clearly derived from these early inscriptions, structurally unchanged but slightly modified, as shown in Table 5.7 (Fossey 1948: 99; Haile 1996: 574).

Table 5.7: Ethiopic numerals (modern)

	1	2	3	4	5	6	7	8	9
1s	፩	፪	፫	፬	፭	፮	፯	፰	፱
10s	፲	፳	፴	፵	፶	፷	፸	፹	፺
100s	፻	፪፻	፫፻	፬፻	፭፻	፮፻	፯፻	፰፻	፱፻
1000s	፲፻	፳፻	፴፻	፵፻	፶፻	፷፻	፸፻	፹፻	፺፻
10,000s	፼	፪፼	፫፼	፬፼	፭፼	፮፼	፯፼	፰፼	፱፼

The signs used in this system are not associated the signs of the Ethiopic script. Instead, they are derived from the letters of the system's Greek or Coptic ancestor. Even though it would not be possible for the signs to have non-numerical meanings, the signs in Table 5.6 have marks both above and below them to indicate that their value is numerical. This practice is universal only from the 15th century onwards, and it is not found at all in the Aksum inscriptions (Ifrah 1998: 246-7). In addition to these signs, modern Amharic texts use an unusual sign for 1000 (ሺ), which lacks marks above and below it (Bender *et al.* 1976: Table I). It may be that the minimal demand for writing very high numbers led to the abandonment of the higher multiplicative formations and the subsequent introduction of an indigenous 1000-sign in Amharic.

The system has a hybrid structure: it is ciphered-additive for the units and tens but multiplicative-additive for numbers over 100. For 10,000, a multiplicative sign consisting of two ligatured 100-signs was used, as shown above: ፼ (10,000) = ፻ x ፻. Thus, because there are special signs for 100 and 10,000, the Ethiopic system has a mixed base of 10 and 100, with the base-10 formations governed by eighteen ciphered characters for 1-9 and 10-90 and the hundreds and ten thousands governed by multiplicative exponent-signs that combine with the ciphered signs. This flexibility allowed the Ethiopic system to express very high numbers with a very limited set of numeral-signs; thus, 647,035 could be written as ፷፬፼፸፻፴፭ (60+4) x 10,000 + 70 x 100 + 30 + 5; and 100,000,000 was simply ፼፼ (10,000 x 10,000). It is unclear whether Guitel's assertion that this system could be extended infinitely by adding additional ፻ signs as needed is accurate, but doing so would not require any new signs or other structural changes (Guitel 1975: 272-3).

It is clear that the Ethiopic system is a member of the alphabetic family. Its development occurred under cultural contact and Christianization by Egyptian and Syrian missionaries. Most extant sources assume that borrowing of the signs must have been from the Greek uncial script (Ifrah 1998: 246; Bender *et al.* 1976: 124). However, the possibility, raised by Haile, that this transmission might have taken place by means of a Coptic intermediary cannot be dismissed (Haile 1996: 574). Far too little paleographic study has been undertaken to resolve this issue one way or the other. The undeniable fact that Egyptian missionaries were active in Ethiopia throughout the 4th century AD might tip the scales slightly in favour of a Coptic origin. St. Frumentius, generally held to be the first major converter of the Aksumites, was a Syrian by birth, trained by Greeks, but his missionary work was based in Alexandria and focused on establishing connections between the Aksumites and the Egyptian Copts. There does not appear to have been any influence on the Ethiopic numerals from the South Arabian script, which was at the end of its lifespan by the time the script was being developed, and for which there is no evidence of the use of numerical notation in the latter part of its history.

The use of base-100 for the multiplicative component of the system is the Ethiopic system's most notable feature. It is the only one of the alphabetic systems to be multiplicative-additive starting at 100; none of the others begins using the multiplicative principle until 1000 or 10,000. This innovation had a clear antecedent in the Greco-Coptic use of multiplication at 10,000, but it eschews the extra 9 signs for 100-900 needed for both Greek and Coptic numerals, and also is more regular than either system in its use of multiplication. The Ethiopic system employs the entire set of 18 ciphered signs, then the same set again beside the 100-sign, then the same set beside the 10,000-sign, whereas both Greek and Coptic use 100-900, then 1 through 9 with a hasta or sub-stroke for the thousands, then start at 1 again for the ten thousands and beyond.

The origins of this unusual structure are poorly understood, but it may be a consequence of the fact that, while the Ethiopic numerals were based on the Greek or Coptic ones, the Ethiopic script was not. Of all the descendants of the Greek alphabetic numerals, the Ethiopic system is the only one to use non-phonetic signs as numeral-signs. In so doing, it becomes more cumbersome, since one needs to learn all the script-signs as well as twenty distinct numeral-signs. On the other hand, if the numeral-signs used are not simply the characters of the script taken in some pre-determined order, there is no impetus to use all the signs, if doing so would require even more effort. In systems that assign numerical values to an ordered series of script-signs, it is natural that one would assign values to all the signs, rather than stopping at some arbitrary point. In the Ethiopic case, the numeral-signs borrowed were different from the signs used for phonemes. Thus, the Aksumites borrowed the first nineteen symbols of the Greek alphabet and then, rather than adopting another nine signs for 100-900 – signs that would have been meaningless to them – decided to take the sign for 100 and use the multiplicative principle thereafter, thereby reducing the number of new signs they needed to learn.

There is extensive evidence on coins (indicating regnal years) and on stone inscriptions (indicating cardinal and ordinal quantities of various kinds) from the Aksumite epigraphic record. However, there is no evidence for the use of the Ethiopic numerals for arithmetic or mathematics; presumably, Ethiopian mathematicians would have used the Greek, Coptic, or Arabic numerals, depending on where they received their training. After the fall of the kingdom of Aksum and the Islamic conquest, the Ethiopian script was used only rarely, and over time became an esoteric script known only to priests and other learned men associated with the Ethiopian Orthodox Church. The numerals are still used for pagination and stichometry in liturgical texts of that church. In Amharic texts, they are used for a wider variety of functions from the 15th century to the present day; for instance, they were found in the personal correspondence of Amharic elites in the 19th century (Pankhurst 1985). The New Testament printed in Amharic in 1852 uses the Ethiopic numerals throughout for page, chapter, and verse numbers (Novum Testamentum in linguam amharicam 1852). Today, the numerals are still occasionally used for writing dates, but have largely been supplanted by the Western numerals (Bender *et al.* 1976: 124).

Gothic

The Gothic alphabet was developed around 350 AD by Wulfila, a bishop who translated the Bible into his native language. Gothic was an East Germanic language spoken by the Germanic tribes who migrated throughout Europe in the latter years of the

Western Roman Empire (Ebbinghaus 1996: 290). The script was alphabetic and written from left to right. Along with the script, an alphabetic numerical notation system was employed, as indicated in Table 5.8 (Braune and Ebbinghaus 1966: 10).

Table 5.8: Gothic numerals

	1	2	3	4	5	6	7	8	9
1s	𐌰	𐌱	𐌲	𐌳	𐌴	𐌵	𐌶	𐌷	𐌸
10s	𐌹	𐌺	𐌻	𐌼	𐌽	𐌾	𐌿	𐍀	𐍁
100s	𐍂	𐍃	𐍄	𐍅	𐍆	𐍇	𐍈	𐍉	𐍊

Like the Greek alphabetic numerals, the Gothic numerals were usually distinguished from the rest of the text either through dots to either side of the numeral-phrase (e.g. •𐍇𐌾𐌴• = 665) or by placing a horizontal stroke above the phrase (Braune and Ebbinghaus 1966: 10). The Gothic numerals were never used to express quantities higher than 1000, and thus there is no evidence of the use of the multiplicative principle. Larger numbers appear to have been always written out in full (Menninger 1969: 260). The system is therefore ciphered-additive and decimal throughout.

The numeral-signs are clearly related to the Greek uncial letters that were used as alphabetic numerals. Of the episemons, 𐌵 (6) was the sixth letter of the Gothic alphabet, and had the phonetic value [k^w], while the other two episemons, qoppa (𐍁) and san (𐍊), had no phonetic value and were simply used to fill out the full complement of 27 signs. The possibility has been considered, but now largely rejected, that the Gothic script owes its ancestry at least in part either to the Latin alphabet or the Germanic runes (Ebbinghaus 1996: 290-291). However, since neither of these other scripts uses alphabetic numerals, the Gothic numerals are clearly of Greek origin.

The Gothic alphabet is attested in only a limited set of documents, mostly translations of parts of the New Testament, but also on a small number of secular texts. Most numerals in Gothic texts therefore serve to indicate chapter and verse numbers in Bibles. Additionally, they were used within the text to indicate numerical values, while in Greek such numbers were always written out in full (Menninger 1969: 260). There is no evidence that the Goths ever did arithmetic or mathematics using these numerals.

Very little of what surely must have been written in Gothic has survived to this day. Most surviving texts date to the 6th century AD, although the assignment of a 4th century AD origin to the numerals is undisputed. It is unclear exactly when in the seventh or early eighth century the Gothic language died out, but around that time the script and numerals ceased to be used, and were replaced by the Roman numerals that were coming to be used throughout Western Europe.

Hebrew alphabetic

The earliest Hebrew scripts began to diverge from the earlier Phoenician consonantary in the 9th or 10th century BC. Then as now, Hebrew consisted of 22 consonantal signs, written from right to left, and placed in a customary order. Early Hebrew inscriptions used a variant of the Egyptian hieratic numerals (ch. 2). Somewhat later, particularly in the 5th and 4th centuries BC, many Hebrew speakers used the Aramaic numerals (ch. 3) for administrative and commercial purposes, as found in the Hebrew Aramaic papyri from Elephantine. Only at a much later date, probably in the 2nd century BC, did a uniquely Hebrew set of numerals develop. This system is indicated in Table 5.9.

Table 5.9: Hebrew alphabetic numerals (Hasmonean)

	1	2	3	4	5	6	7	8	9
1s	א	ב	ג	ד	ה	ו	ז	ח	ט
10s	י	כ	ל	מ	נ	ס	ע	פ	צ
100s	ק	ר	ש	ת	תק	תר	תש	תת	תתק

The first 22 signs, indicating 1 through 400, are the letters of the Hasmonean Hebrew script as it was used about 125 BC, at the time of the writing of the Dead Sea

Scrolls (Goerwitz 1996: 488). I present this particular script because Hebrew alphabetic numerals are first encountered on inscriptions from the Hasmonean Dynasty. The system is decimal and ciphered-additive, and written from right to left; thus, 369 would be written שסט. Values between 500 and 900 were represented using the sign for 400 in conjunction with one or more signs for the lower hundreds (i.e. 500 = 400+100, 600=400+200 ... 900 = 400+400+100). This structural irregularity exists because there are too few letters in the Hebrew consonantary to fill out the twenty-seven signs needed to extend the system to 900. This irregularity does not change the fact that the system is ciphered-additive and decimal, but the sign 400 occupies a special structural role. 400 is not a base of the system, however, as its exponents (16,000, 6,400,000, etc.) do not receive any special treatment.

While the very earliest Hebrew inscriptions contain no signs for numbers above 1000, the need to do so quickly arose, as the numerals began to be used for dating on grave inscriptions using the Hebrew calendar. For multiples of 1000, a mark - either a small curved stroke to the left of a numeral-sign or two dots placed above it - could be used to indicate multiplication by 1000; thus, ט׳ would signify 9000 and צ׳ 90,000. This feature is similar to, but distinct from, the Greek alphabetic system, which adds a stroke above or below a numeral to indicate multiplication by 1000, but begins again at 10,000 by placing the multiplicand above the sign M. Thus, while the Greek system could express any numeral up to 10 million, as opposed to one million for the Hebrew numerals, the Hebrew system is arguably easier learn and use by virtue of only having one value at which the multiplicative principle is employed.

Claims of a very early origin (9th to 7th century BC) of the use of the alphabetic numerals have now been thoroughly discredited. The primary evidence in favour of this position was the assumption that since the Greek alphabet had a Semitic (Phoenician) origin, the Greek alphabetic numerals must also have had an early date and Semitic origin (Gandz 1933: 75-6; Schanzlin 1934; Smith and Karpinski 1911: 33). Zabilka's musings on the possible early origin of the system rest on a confusion of the hieratic and Aramaic numerals mentioned in Chapter 2 and 3, respectively, with the alphabetic numerals, and cannot be taken seriously (Zabilka 1968: 176-78). There is no evidence for the use of the Hebrew script numerically before the 2nd century BC, while there is strong negative evidence supporting the hypothesis of a late origin. For instance, on the Khirbet el-Kôm ostracon, a bilingual inscription in Aramaic and Greek found in Palestine and dating to 277 BC, the Aramaic portion of the inscription uses Aramaic numerals while the Greek portion uses Greek numerals (Geraty 1975). If the Semites of the Levant were using the Hebrew alphabetic numerals at that time, they likely would have used them instead of the Aramaic numerals in a situation where the Greek inscription a few lines below used a similar system.

It seems that the Hebrew numerals were used numerically in a very limited sense at an early date. Rather than being a full numerical notation system, however, this system was used to label a limited set of entities by means of the letters in order from 1 to 22, without a ciphered-additive or decimal structure. This system is called the *alphabetic ordinalia* by Gandz (1933: 77). In such a system, aleph = 1, beth = 2, gimel = 3, but beyond 10, the ordering simply continues by ones, until reaching shin = 21 and taw = 22. This is probably the system used on coins minted at Sidon dating from the Alexandrine period, but because the attested numbers are all lower than 10, it may be that they were part of a full ciphered-additive system for which evidence of higher numbers has been lost (Harris 1936: 19). This "letter-labelling" was used at a very early date in Greece, but does not represent the use of an alphabetic or any other structured numerical notation system. In either the Greek or Hebrew inscriptions, if there are no numerals above 10, the alphabetic ordinalia cannot be distinguished from the full-fledged ciphered-additive system, but we must not surmise the existence of the latter from the existence of the former.

The best evidence we now have suggests that the Hebrew alphabetic numerals were first borrowed from the Greek alphabetic numerals between 125 and 100 BC for use in coins inscribed with the Hasmonean script developed in the 2nd century BC (Gandz 1933: 76; Millard 1995: 192). The first safely dated instance on which Hebrew alphabetic numerals are certain is on coins from the reign of the Hasmonean king Alexander Janneus (103 to 76 BC), some of which were stamped with Greek script and numerals, others with the Hebrew script and numerals, in both cases using alphabetic signs as described above (Naveh 1968). Yet there is a clay seal upon which the inscription "Jonathan high priest Jerusalem M" was found, Avigad believes this individual to be Alexander Janneus, whose Hebrew name was Jonathan, thus placing the seal in the same period (Avigad 1975: 10). Avigad suggests that the M (=40) might signify the fortieth year of the chronology established when the Hasmonean kings took power in 142 BC, meaning that the inscription would date to 103 BC (Avigad 1975: 11-12). That these early Hebrew alphabetic numerals would both be found in the context of the same man is surely suggestive, though not conclusive. Regardless of the specific context of their development, a late 2nd or early 1st century BC origin for the numerals – the period of the Hasmonean kings - is generally accepted today.

Aside from the special formation of numerals from 500 to 900, the similarities between the Greek and Hebrew numerical notation systems are striking. The two systems share not only a similar structure (decimal and ciphered-additive) but also a similar principle for forming the numeral-signs (alphabetic). The notion that the Hebrew numerals were independently developed can no longer seriously be sustained, despite the agnostic attitude of some scholars, including Ifrah (1998: 239). That Hasmonean coins were struck in both languages and using both systems provides specific contextual evidence that the Hasmonean kings adopted the technique from the Greeks. The cultural influence of the Ptolemaic and Seleucid kingdoms in the Levant at this time was enormous; Greek alphabetic numerals were used on coins from the Phoenician cities of Sidon, Tyre, Byblos, and Akon from the mid-3rd century BC onward (Millard 1995: 193). In contrast to the Aramaic numerals previously in use, which were cumulative-additive for values below 100 and thus required many signs to represent even small numbers, the alphabetic numerals were very concise and thus well suited for short inscriptions on coins. Given this confluence of different lines of evidence, it is evident that the model for the Hebrew alphabetic numerals was the Greek alphabetic system. At the same time, the Hebrew use of 400 as a 'stepping-stone' for representing the higher hundreds is an important innovation, as it did not require that Hebrew speakers learn and adopt additional non-phonetic signs. This development did not occur in Greece, where it was necessary to borrow the episemons from archaic Greek and Semitic scripts.

The use of the Hebrew script for the numerals, rather than borrowing the Greek numerals wholesale, represents the earliest development of a distinctively Hebrew system. This is particularly notable because most Semitic peoples of the ancient Levant (Nabataeans, Aramaeans, Palmyrans) never used alphabetic numerals, but continued to use their own hybrid cumulative-additive / multiplicative-additive systems (ch. 3). These systems coexisted with the Hebrew alphabetic system for several centuries, and were only replaced over a long period. In the early history of the Hebrew alphabetic numerals (up to the 7th century AD), inscriptions on Jewish graves throughout the Mediterranean region were often written, not with the Hebrew numerals, but rather in the Greek alphabetic numerals with which the carvers were also familiar (Ifrah 1998: 238-9). This is further evidence in favour of a Greek origin of the alphabetic numerals, for it seems unlikely that, if the Greeks borrowed the numerals from the Hebrew script, Jewish carvers would borrow them back for grave inscriptions. This early period was characterized by the slow displacement of the Greek and Levantine systems by the Hebrew numerals, until by the Middle Ages they were firmly established as a distinctive system peculiar to the Jewish populations of Europe, North Africa, and the Levant. Also, in the 6th century AD, the Hebrew numerals were partly or wholly used by the creators of the Syriac *estrangelo* alphabetic numerals (see below), which are also ciphered-additive and decimal and have the same break at 400 as the Hebrew system.

The classical and modern Hebrew alphabetic numerical notation system has the same structure as the ancient system, but uses modern script-signs, as shown in Table 5.10.

Table 5.10: Hebrew alphabetic numerals (modern)

	1	2	3	4	5	6	7	8	9
1s	א	ב	ג	ד	ה	ו	ז	ח	ט
10s	י	כ	ל	מ	נ	ס	ע	פ	צ
100s	ק	ר	ש	ת	תק	תר	תש	תת	תתק

Around the beginning of the 10th century, the option arose of using five additional Hebrew characters to complete the sign set for 500 through 900. These signs (ך ם ן ף ץ) are the signs used for *kof, mem, nun, pe,* and *tsade* when those characters are in word-final position. These forms were used in some of the Masoretic commentaries on the Old Testament, but do not appear to have ever been the regular forms used for the numerals (Gandz 1933: 96-102). They were certainly not the common forms used in the 12th century, as Ifrah has shown from numerous Jewish gravestones in Spain (Ifrah 1998: 216). Today, the older formations using additive combinations of hundreds-signs are the sole means of expressing values above 500 in the alphabetic numerals. Gandz asserts that the main reason these forms did not become widely accepted was that the word-finality of these signs was inconsistent with the principle of the numerical notation system that the highest values should come first, rather than last, in a numeral-phrase (Gandz 1933: 98). To put a word-final letter for 500-900 at the head of a numeral-phrase would have been inconsistent with its original purpose; to put it at the end of the numeral-phrase would be inconsistent with the rule of decreasing sequential ordering of the exponents.

Also in the 10th century, Hebrew scholars became aware of the Hindu and Arabic positional numerals, and occasionally experimented using combinations of the alphabetic numerals and the positional principle. In a Masoretic poem by Saadia Gaon (882-942

AD), numbers are written in two positional columns; the rightmost represents the thousands position, while the leftmost column represents the ones, so that 42,377 was written as שעז   מב (Gandz 1970: 487-488). Starting in the 12th century, some Hebrew writers simply used the first nine alphabetic symbols in place of the ordinary Western or Arabic signs, supplementing them with a circle for zero, and thus converted their system into a fully ciphered-positional one (Gandz 1933: 110). This practice was apparently first used by Abraham Ben-Ezra in his *Sefer Hammispar* (Book of Number) written about 1160, although the ordinary Hebrew numerals were always used in the regular text of such works, with the positional variants used only for mathematics (Schub 1932). This technique was never commonly used, however, and most later Hebrew mathematicians and astronomers simply used Western or Arabic numerals.

One of the more important functions for which the Hebrew numerals have been used historically is *gematria*, the art of number-magic (Ifrah 1998: 250-256). Because every letter of the Hebrew script has a numerical value, every Hebrew word has a numerical value equal to the sum of its letters' values. Among medieval and early modern scholars, this practice was commonly employed for interpreting passages from the Talmud and the Midrash and for finding symbolic associations among words that share the same numerical value. For instance, two of the terms associated with the Messiah, *shema* צמח 'seed' and *menakhem* מנחם 'consoler', have the same numerical values (8+40+90 = 8+40+50+40 = 138). A related practice is the construction of *chronograms*. A chronogram is a verse in which a specified set of words has a numerical value equal to the date of an event (e.g. a person's death) to which the entire verse refers. Because these practices (also used with the Arabic abjad, described below) can only be done where a system exists for correlating phonetic signs with numerical values, they probably contributed to the continued use of the corresponding numerical notation systems long after ciphered-positional systems had been adopted for most purposes.

While the Western numerals are used in modern Israel for most purposes, the alphabetic system is regularly used for dates using the traditional Jewish calendar, especially in religious texts and on graves. In 1999, the Israeli Supreme Court ruled that gravestones in orthodox Jewish cemeteries could henceforth record numbers using Western numerals rather than the Hebrew alphabetic system, whose use had previously been mandatory in that context (Copans 1999). It is unclear whether this ruling will have any long-term effect on the use of the Hebrew numerals for dating. However, the fact that such a ruling needed to be made at all shows the continued health and vibrancy of Hebrew alphabetic numerals, albeit in a limited set of religious contexts. The Hebrew alphabetic numerical notation system is now over two millennia old, and is one of the oldest systems in continuous and regular use.

Syriac alphabetic

In Chapter 3, I described the numerals used alongside the Syriac Estrangelo script used between about 50 and 500 AD. This hybrid cumulative-additive / multiplicative-additive system was related to the others used in the region, such as Aramaic, but by the 5th and 6th centuries AD most of these other systems had ceased to be used. Around this time, the Estrangelo script diverged into two forms: an eastern variety, Nestorian, used by the Christians of Persia, and a western variety, the Serto script used by the Jacobite Christians of Syria. This split was precipitated by the expulsion of the Nestorian Christians from the Byzantine city of Edessa and their subsequent migration into the Sassanian Empire (Duval 1881: vii). Soon thereafter, both the Nestorian and Serto scripts began to use an alphabetic numeral system akin to those used elsewhere in the Middle East. The basic signs of this system (as used in the Serto script) are shown in Table 5.11.

Table 5.11: Syriac alphabetic numerals

	1	2	3	4	5	6	7	8	9
1s	ܐ	ܒ	ܓ	ܕ	ܗ	ܘ	ܙ	ܚ	ܛ
10s	ܝ	ܟ	ܠ	ܡ	ܢ	ܣ	ܥ	ܦ	ܨ
100s	ܩ	ܪ	ܫ	ܬ	ܬܩ	ܬܪ	ܬܫ	ܬܬ	ܬܬܩ
1000s	ܐ̗	ܒ̗	ܓ̗	ܕ̗	ܗ̗	ܘ̗	ܙ̗	ܚ̗	ܛ̗
10,000s	ܐ̱	ܒ̱	ܓ̱	ܕ̱	ܗ̱	ܘ̱	ܙ̱	ܚ̱	ܛ̱

The system is decimal and ciphered-additive, and written from right to left, so that 369 would be written as ܫܣܛ. As in many alphabetic numeral systems, numerals were sometimes distinguished from the rest of the script by placing a horizontal stroke above a numeral-phrase, but often no special mark was present. Like the Hebrew alphabetic numerals, values from 500 to 900 were usually expressed using the signs for the lower hundreds with the sign for 400 in various additive combinations. Alternately, the upper hundreds were occasionally expressed multiplicatively, by placing a small dot above the signs for 50 through 90 to indicate multiplication by 10: ܢ̇ ܣ̇ ܥ̇ ܦ̇ ܨ̇ (Duval 1881: 15; Noldeke 1904: 316-317). For values above 1000, the multiplicative principle was always used, a slanted stroke placed beneath a unit-sign indicates multiplication by 1000, while a horizontal stroke placed beneath a sign indicates multiplication by 10,000 (Ifrah 1998: 240). In this way, any number below 10 million could be expressed. Duval (1881: 15-16) claimed that even higher values could be expressed by placing two small strokes beneath a sign to indicate multiplication by ten million, and that placing one small stroke above and one small stroke beneath a sign indicated multiplication by ten billion; however, these techniques were used extremely rarely. As in several other alphabetic systems, placing an oblique stroke above a given numeral-sign indicated the appropriate (1/x) unit fraction (Duval 1881: 16).

The origin of the Syriac alphabetic numerical notation system remains enigmatic, as both its date and ancestor are poorly known. It appears to have been invented in the late 6th or early 7th century AD, as there is no surviving textual evidence for its use before that time (Duval 1881: 15). It was not accepted instantaneously. A manuscript dated from the 7th or 8th century (British Museum Add. 14 603) is paginated both in the Syriac alphabetic numerals and in the older Estrangelo system (Ifrah 1998: 241). This suggests that the alphabetic system was beginning to supplant the earlier means of expression but was not yet fully adopted or understood.

Independent invention of this system can be ruled out quickly, given its strong similarity to others used in the region. Two likely possibilities are that it was modelled on the Greek alphabetic numerals prevalent in the Byzantine Empire or else on the Hebrew alphabetic numerals. Of these, the latter theory perhaps has the most to recommend it. The Hebrew and Syriac numerals are the only two systems in which 400 occupies a special structuring role, in that the higher hundreds are expressed using additive combinations of 400 and the lower hundreds. The ordering of the Syriac numerals follows the letter-order shared by the Syriac and Hebrew scripts. If Syriac had developed from a Greek model, we might expect to see the numerals valued according to a Greek-derived order, or perhaps extra signs would have been borrowed to provide the five extra signs needed to complete three sets of nine signs. Finally, the Syriac system was written from right to left like the Hebrew system, but unlike the Greek.

On the other hand, the Syrians were closely affiliated with Eastern (Greek) Christianity, and many Syrians lived under Byzantine rule, so it is possible that the Syriac numerals derive from a Greek rather than a Hebrew ancestor. The hypothesis of Greek ancestry is supported by the shared feature of the two systems that the multiplicative principle was used at two different exponents of the base, 1000 and 10,000, whereas the Hebrew numerals only did so at 1000. It is probable that the inventor(s) of the Syriac numerals would have been familiar with both the Greek and Hebrew numerals. It is thus possible that features of both systems were blended in the development of the Syriac system. Of the two, the case for Hebrew ancestry seems stronger, if only because of the unusual feature of special structuring after 400, whereas the Greek influence may have been more indirect.

The Syriac scripts were never used as the official script of any polity, and thus Syriac numerals are rarely found on stone monuments or coins. However, their use in religious texts is extremely prevalent from the time of their invention to the present day, a situation afforded by the relative separation of the Jacobites and Nestorians from both Western and Eastern Christianity. Most Syriac manuscripts are dated and paginated using the numerals, making it easy to examine paleographic changes in the numerals over time. There is no evidence that the Syriac numerals were ever used for mathematics, nor, in contrast to the Hebrew numerals, were they used for letter-magic or numerology.

Although the heyday of the Syriac scripts came and went before the year 1000, both the Nestorian and Serto scripts survive to this day, the former in Iraq, Turkey, and Iran among a small number of Nestorian Christians, the latter in Lebanon among the Maronite Christians of that country. Both scripts have retained their distinctive numerical notation systems to the present day. Yet these systems are greatly restricted in the contexts of their use to the same liturgical functions for which they have been employed for nearly 1500 years. Arabic and/or Western numerals are used for most other purposes. Nevertheless, there is no reason to think that the Syriac numerals are about to disappear, particularly given the special status accorded to Maronite Christianity in Lebanon's 1990 constitution.

Arabic abjad

Before the rise of Islam, Arabic speakers used a variety of the Nabataean script and numerals (ch. 3). While the classical Arabic script is directly descended from this ancestral form, the Nabataean hybrid cumulative-additive / multiplicative-additive numerals were abandoned in favour of a ciphered-additive system based on the Arabic script-signs. The basic signs of this system are shown in Table 5.12 (Saidan 1996: 332).

Table 5.12: Arabic abjad numerals

	1	2	3	4	5	6	7	8	9
1s	ا	ب	ج	د	ه	و	ز	ح	ط
10s	ي	ك	ل	م	ن	س	ع	ف	ص
100s	ق	ر	ش	ت	ث	خ	ذ	ض	ظ
1000	غ								

The system is decimal and ciphered-additive, and, like the Arabic script, is written from right to left with the signs in descending order. The numeral-signs shown are the unligatured signs of the Arabic consonantary; in numeral-phrases, signs were ligatured to one another as appropriate for the letters in question. A horizontal stroke sometimes was placed above a numeral-phrase to distinguish it from an ordinary word. Curiously - and importantly for understanding the history of the system – the signs are not valued according to the normal Arabic letter-order, but rather according to the letter-order of the Hebrew and Syriac scripts, which was also used by the Arabs early in their script's history. Because its first three signs (*'alif, ba, jim*) are the first three of the Hebrew script, the system was sometimes called *hisab abjad*, from which the name assigned it in modern scholarship is derived.[3] Because the Arabic script has 28 basic consonantal signs, the remaining sign, *ghayin* (غ), was assigned the numerical value of 1000. The greatest importance of this sign was not as a single unit-sign, but as part of a larger multiplicative structure that was used for values beyond 1000 by placing another sign before the 1000-sign. Thus, 7642 might be written (from right to left) as زغخمب (7 x 1000 + 600 + 40 + 2), although with individual signs ligatured together. In this way, any number up to and

[3] The first three letters in the modern Arabic script are *'alif, ba,* and *ta.*

including one million could be written, more than sufficient for the needs of classical Arabic civilization.

To further complicate matters, the abjad numerals used throughout most of the Islamic world were modified somewhat among users of the Arabic script in North Africa and Spain, in that the values assigned to six signs were changed. This ordering developed somewhat later than that used in the east, perhaps in the 9th century AD. Other than the different values assigned to the six signs in Table 5.13 below, the system is structurally identical to the regular abjad numerals.

Table 5.13: Arabic abjad numerals (Eastern vs. Maghreb)

Sign	Letter-name	Eastern value	Maghreb value
س	sin	60	300
ص	sad	90	60
ش	shin	300	1000
ض	dad	800	90
ظ	dha	900	800
غ	ghayin	1000	900

Given the importance of the Arabs in the period between 600 and 1000 AD, it is surprising that so little attention has been paid to the origins and early history of the abjad numerals used throughout this period. The exact date of their origin is still unknown, although it is certain that the system developed in the 7th century AD. It is remotely possible that the system is of pre-Islamic origin, and that it spread from the north. However, it is more likely to have originated around 650 AD, at or shortly after the time of the early Islamic conquests in Syria, Egypt, and Mesopotamia. Under Byzantine rule, this region had used the Greek alphabetic numerals (along with Roman numerals) for administrative and commercial functions; furthermore, both the Syriac and Hebrew alphabetic numerals were used in their respective scripts. Thus, the independent invention of the abjad numerals is highly implausible, given their similarity in structure and letter-order to these three other systems. The question then becomes which of these three systems is the immediate ancestor of the abjad numerals, or whether more than one system was used as their model. Given that all three systems were used regularly in the 7th century AD in the Middle East, structural features and considerations of historical context must be invoked to determine which ancestors were most important.

Although the Arab script today uses its own letter-order for its consonantary, its numeral-signs have exactly the same order as the corresponding Syriac and Hebrew characters up to 400 (above which point the other two systems are structurally irregular). This is because the earlier Arabic script used the North Semitic letter-order when the numerals were invented, and though the script's letter-order was altered in the 8th century AD, the older order was retained for the numerals. The Greek system follows a similar order to the Syriac and Hebrew numerals up to the 80, but diverges thereafter by putting the third episemon, sampi (equivalent to the Hebrew tsade) at the end of the system, rather than in the middle. This demonstrates that the Hebrew or Syriac systems probably played a significant role in the development of the Arabic abjad numerals (Guitel 1975: 276-278; Ifrah 1998: 243). Table 5.14 illustrates how the Arabic order is directly parallel to the Hebrew and Syriac, while the Greek numerals diverge from them starting at 90.

Table 5.14: Arabic, Greek, Hebrew, and Syriac numeral-signs and letters

	Arabic		Greek		Hebrew		Syriac	
1	ا	'alif	Α	alpha	א	aleph	ܐ	olap
2	ب	ba	Β	beta	ב	bet	ܒ	bet
3	ج	jim	Γ	gamma	ג	gimmel	ܓ	gomal
4	د	dal	Δ	delta	ד	dalet	ܕ	dolat
5	ه	ha	Ε	epsilon	ה	he	ܗ	he
6	و	wa	Ϝ	vau	ו	vov	ܘ	waw
7	ز	zay	Ζ	zeta	ז	zayin	ܙ	zayn
8	ح	ha	Η	eta	ח	het	ܚ	het
9	ط	ta	Θ	theta	ט	tet	ܛ	tet
10	ي	ya	Ι	iota	י	yod	ܝ	yud
20	ك	kaf	Κ	kappa	כ	kof	ܟ	kop

	Arabic		Greek		Hebrew		Syriac	
30	ل	lam	Λ	lambda	ל	lamed	ܠ	lomad
40	م	mim	Μ	mu	מ	mem	ܡ	mim
50	ن	nun	Ν	nu	נ	nun	ܢ	nun
60	س	sin	Ξ	xi	ס	samekh	ܣ	semkat
70	ع	ayin	Ο	omicron	ע	ayin	ܥ	'e
80	ف	fa	Π	pi	פ	pe	ܦ	pe
90	ص	**sad**	ϙ	**qoppa**	צ	**tsade**	ܨ	**sode**
100	ق	qaf	Ρ	rho	ק	quf	ܩ	quf
200	ر	ra	Σ	sigma	ר	resh	ܪ	rish
300	ش	shin	Τ	tau	ש	shin	ܫ	shin
400	ت	ta	Υ	upsilon	ת	tav	ܬ	taẘ
500	ث	tha	Φ	phi	ך		ܢ̇	
600	خ	kha	Χ	chi	ם		ܣ̇	
700	ذ	dhal	Ψ	psi	ן		ܥ̇	
800	ض	dad	Ω	omega	ף		ܦ̇	
900	ظ	dha	ϡ	sampi	ץ		ܨ̇	

On the other hand, the Arabic abjad was rarely used in the same texts as were either the Hebrew or the Syriac numerals. In contrast, Greek alphabetic numerals are found in Arabic documents from the 7th to the 9th centuries AD, and in many texts co-occur with the abjad numerals. In 706 AD, Caliph Walid I dictated that, although his Greek financial administrators in Damascus were no longer to use the Greek alphabet, they would be permitted to continue to use the Greek alphabetical numerals (Menninger 1969: 410). In an 8th century AD Arabic tax record, numbers are expressed in both Greek alphabetic numerals and abjad numerals (Cajori 1929: 29). Because many of the regions conquered by the Arabs – even those such as Syria and Palestine in which the Syriac or Hebrew numerals were found – were under Greek rule, Greek numerals were the normal system used for administration, on coins, and in inscriptions. It is unlikely that the Greek system played no role in the development of the abjad numerals.

However, Ifrah's suggestion that the Arabs followed the Greek example for the final six letters of their numeral system (500 through 1000) is incomplete (Ifrah 1998: 243). The phonetic values of the final six Arabic characters do not correspond with the Greek. Furthermore, the Greek system has only twenty-seven rather than twenty-eight signs lacking a sign for 1000). While the Arabs were no doubt aware that the Greek system had signs for the higher hundreds, and may thus have attached numerical values to the remaining signs in their own script, the use of a special sign for 1000 is unique to the Arabic system among all four systems under consideration. More likely, the Arabic system was based on the Semitic letter-order but employed the structural advantages of a system, such as the Greek, with a full complement of numeral-signs. This feature would have been particularly important, since the administrative needs of the new Islamic caliphate were growing exponentially.

By the late 8th century AD, the Arabic abjad numerals had spread throughout the Middle East and into the Maghreb. They were used on administrative documents, in literary and scientific texts, and on monuments, though not for the most part on ephemeral media such as ostraca. In areas in which an existing administrative apparatus was retained from the Byzantine Empire (such as Egypt), the Greek and Coptic alphabetic numerals were used much more frequently on administrative and financial documents than were the abjad numerals (Grohmann 1952: 89). Because these systems were structurally similar to the abjad numerals, consideration of their utility for specific functions is irrelevant. Yet it would have taken some effort for Arabic writers to learn an entirely new set of 27 signs, so its failure to be used more widely is somewhat surprising. Issues of identity and ethnicity may have played a significant role in determining the scope of their use. For instance, Ifrah describes a 9th century AD Christian manuscript written in Arabic but in which the verses are numbered in Greek (Ifrah 1998: 243). In this case, it is probably because the writer was a Christian who associated himself with Greek Christianity through the alphabetic numerals even though he wrote in the Arabic language and script.

Despite their use over a wide area, the abjad numerals did not give rise to a large number of descendant systems. In part, this must be due to the remarkable stability of the Arabic script itself. The Coptic "Epakt numerals" used in Egypt from the 10th century

AD onward (for which see the Coptic numerals above) are an interesting blend of Greek and Arabic influences used in some administrative documents. Though some of the Epakt numeral-signs are based on those of the abjad, it is unclear to what extent the Arabic abjad affected the structure of the system, since the classical Coptic numerals are decimal and ciphered-additive. A similar situation probably arose in Morocco, where "Fez numerals", incorporating elements of the Arabic abjad as well as Greco-Coptic alphabetic numerals, were used until very recently (Colin 1933). Finally, Arabic astronomers used a very unusual system for writing fractions, which combines a ciphered-additive system with a base-60 (sexagesimal) positional notation. This system was based on a similar Greek astronomical system that used the Greek alphabetic numerals.

Shortly after they were invented, the abjad numerals began to be supplanted by another, more famous numerical notation system – the ciphered-positional Arabic system (ch. 6) borrowed from the Hindus. Attié Attié (1975) contends that the ciphered-positional numerals were developed in pre-Islamic Arabia by 568 AD, but this conclusion is based on a misunderstanding of a bilingual inscription and must be discounted. Yet the Islamic conquest of enormous territories to the east brought the Arabic and Indian spheres of influence into close contact by the mid-seventh century. In a text dated to 662 AD, Severus Sebokht, the Syrian Christian bishop at the monastery at Kenneshre, wrote in admiration concerning a Hindu technique of numeration making use of nine signs (Nau 1910). While Sebokht mentions neither zero nor the positional principle, he probably was referring to the Indian positional numerals, since the ciphered-additive systems of India, like all ciphered-additive systems, require far more than nine signs.

As Islam spread eastward throughout the 8th century AD as far as the Indus River, the Indian style of numeration began to move westward and supplant the Arabic abjad, which itself was still a novelty in western regions such as North Africa. This replacement was greatly hastened by the arrival in 773 AD of Hindu astronomers and astronomical knowledge at the court of Caliph al-Mansur in Baghdad, which formed the basis for the early 9th century AD writings of the famed Arab mathematician al-Khwarizmi, who popularized the Arabic positional system (Menninger 1969: 410-411). By the late 9th century, the positional system was being used in administrative and financial documents, and by the late 10th century, on inscriptions (Grohmann 1952: 89). The latest Arabic papyrus in which abjad numerals are used to write a date is from 517 AH / 1123/4 AD (Destombes 1987: 131).

Astronomers continued to use ordinary abjad numerals much later than other writers, probably because they had also adapted them for use in the quasi-positional sexagesimal fractions described below. Destombes (1987) notes that abjad numerals were commonly used on most Arabic astrolabes (both for marking gradations on the instrument and dates of construction) until the 16th century. As well, the abjad numerals survive to the present for very limited cryptographic, literary, and magical functions. In many texts, dates were concealed in a verse worded so that the total numerical value of its letters in the abjad system yielded the desired date. The technique of writing such concealed dates, or *chronograms*, was known as *hisab al-djummal* (Colin 1971: 468). Chronograms were also common in medieval Hebrew writings (see above). Chronograms using the abjad numerals were common throughout the Middle Ages, particularly in Persia and Islamized parts of India (Ahmad 1973). Finally, the abjad numerals were retained for pagination of prefaces and tables of contents of books, similar to the Western conventional use of the Roman numerals (Colin 1960: 97). Abjad numerals survived particularly well in the Maghreb, and continue to be used there for chronograms, cryptographic correspondence, and functions related to magic and divination. In Morocco, chronograms were very commonly used in the 17th and 18th centuries, but apparently no earlier (Ifrah 1998: 252). Abjad-derived systems were used by 19th century Ottoman administrators for cryptographic purposes (Decourdemanche 1899). Similarly, Monteil (1951) discusses a text from Mali that discusses many cryptographic systems derived from the abjad numerals that were used in North Africa in the mid-twentieth century.[4] It is unclear to what extent these systems are still used today. Most modern Arabic grammars mention the existence of abjad numerals and note the numerical values of the various letters.

Astronomical fractions

For most purposes, the Greeks, Hebrews, and Arabs used the decimal, ciphered-additive alphabetic numerals peculiar to their civilizations. Yet, in many astronomical texts, a distinct set of systems was used for computing and recording fractions. These systems blend ordinary ciphered-additive numerals of the sort discussed above with a base of 60 and the positional principle, as used in Babylonian astronomy. This ingenious representational technique, which I will call astronomical fractions, represents a very curious digression in the history of numerical notation.

In Chapter 7, I will discuss the sexagesimal notation used from about 1800 BC to BC/AD in Babylonian astronomical and mathematical texts, but a word about it here is necessary. This system was cumulative-positional and had a base of 60 with a sub-base of 10. Numbers less than 60 were expressed through cumulative combinations of signs for 10 (𒌋) and 1 (𒁹). For numbers greater than 60, the positional principle was used to express multiples of exponents of 60 (60, 3600, 216000, etc.). Thus, 481042 could be expressed as 𒈫 𒌋𒐈 𒌍𒐛 𒎙𒈫 (2 x 216000 + 13 x 3600 + 37 x 60 + 22).

By the 3rd century BC, Greeks firmly controlled most of the lands formerly under Babylonian rule, under the potent Seleucid kingdom that came into existence after the Alexandrine conquests. In the 2nd century BC, the Babylonian positional notation and the sexagesimal base were married to the Greek ciphered-additive numerals and used thereafter by Greek astronomers (Ifrah 1998: 156). The first major text in which this new

---

[4] Curiously, one of these systems is known as *el-Yunani* (Ionian!), suggesting that its users were aware of the Greek origin of such notations.

system appears was the *Syntaxis* of Ptolemy, written in the 2nd century AD (Heath 1921: 44-45). In place of the cumulative Babylonian signs in each position, fourteen of the Greek alphabetic numerals were used (the units 1 through 9 and the decades 10 through 50), to write any number from 1 through 59. Unlike the Babylonian system, however, the Greek sexagesimal-positional system was never used for expressing integers. Numbers greater than 60 were always written with the ordinary alphabetic numerals. In Greek astronomy, as in modern astronomy, the circle was divided into 360 degrees, each degree into 60 minutes, each minute into 60 seconds, and so on. This system thus corresponds to the various subdivisions of the circle. Thus, in Theon of Alexandria's (4th c. AD) commentary on Ptolemy's *Syntaxis*, the numeral-phrase ͵$\overline{\alpha\phi\iota\varepsilon}$ $\overline{\kappa}$ $\overline{\iota\varepsilon}$ expresses 1515 (͵$\overline{\alpha\phi\iota\varepsilon}$) degrees, 20 ($\overline{\kappa}$) minutes, and 15 ($\overline{\iota\varepsilon}$) seconds (Thomas 1962: 50-51). The degrees value is in the ordinary decimal alphabetic numerals[5], including the use of the multiplicative *hasta* for 1000, while the latter two positions are written in sexagesimal fractions. Sexagesimal fractions did not simply express minutes and seconds, but could be used to express any fractional value. In this sense, the successive positions represent $1/60$, $1/60^2$, $1/60^3$, and so on. Because the system is positional, it can be infinitely extended to express as small a value as desired.

The structure of this system is typologically unusual. Within each position, decimal ciphered-additive numeral-phrases for 1-59 appropriate to the relevant system (Greek, Arabic, or Hebrew) are used. However, the primary base of the system – the one involved in the positional aspect of the system - is 60; thus, 10 (the base of the ordinary alphabetic numerals) becomes the *sub-base* of the sexagesimal fractions. The system's base and sub-base thus follow the Babylonian system, but the intraexponential principle used in forming the signs of each position is ciphered rather than cumulative. This system is thus both ciphered and positional, but it is clearly not identical to ordinary

---

[5] If this value were expressed sexagesimally, we would expect it to be written as $\overline{\kappa\varepsilon}$ $\overline{\iota\varepsilon}$, or (25x60)+15.

ciphered-positional systems such as the Western numerals. Rather, because it has a sub-base, it is interexponentially *ciphered-additive* rather than simply ciphered. It is thus a (ciphered-additive)-positional system.

The astronomical numerals used a special sign for zero as a placeholder to indicate an empty position in order to avoid ambiguity. In some late (4th to 1st century BC) Babylonian texts, a similar placeholder was used to avoid confusion and misreadings. The Greeks adopted this technique using their own sign, which took the form in early manuscripts (1st century AD), and which in later manuscripts was written as (Irani 1955). The latter sign is sometimes held to represent omicron, the first letter of the Greek word *ouden*, "nothing", with a stroke added above to distinguish it from the appropriate letter (Ifrah 1998: 549). Yet it is unlikely that the Greeks would have chosen a sign that already had a numerical value ($\overline{O}$ = omicron = 70) in the alphabetic system (Neugebauer 1957: 14). The zero-sign was probably a paleographic outgrowth of the earlier form, which was, like the Babylonian placeholder sign, purely abstract.

The Arabs, who inherited the bulk of Greek astronomical knowledge when they took control of Mesopotamia in the mid-seventh century, began using astronomical numerals simultaneously with their adoption of the abjad for general numeration. Although using the signs of the Arabic abjad rather than the Greek alphabetic numeral-signs, there can be little doubt, from the date and context of their first use, that the Arabs adopted the astronomical numerals directly from the Greeks, from whom they derived much of their astronomical knowledge. The Arabic sexagesimal numerals are written from left to right, in contrast with systems such as the abjad numerals (Irani 1955: 2). The use of a symbol for zero (, or later) that is derived from the Greek symbols demonstrates that the Arabic system was based on the Greek one (Irani 1955). Similarly, medieval Hebrew astronomers adopted sexagesimal fractions for translations of Arabic and Greek astronomical texts and for their own astronomical calculations, as in the 14th century astronomical writings of Levi Ben Gerson (Ifrah 1998: 158). The cultural history of the alphabetic and astronomical numerals thus involves two separate lines of descent. The first line is that of the regular alphabetic numerical notation systems - the Greek, Hebrew, and Arabic decimal, ciphered-additive systems described earlier in this chapter. The other line stems from the fusion of the Greek and Babylonian systems into the sexagesimal fractions; the resulting "Greek astronomical" system is the ancestor of both the Arabic and Hebrew astronomical systems.

The astronomical fractions were used by almost all astronomers in ancient and medieval Europe and the Middle East from the 2nd century BC until at least the 14th century AD. The use of the Greek variant waned after the end of the Byzantine Empire in the mid-15th century, after which Western or Arabic positional numerals were used for most purposes. The Arabic version of the system survived even longer; Irani (1955: 3) lists many texts from the 16th and 17th centuries and one from as late as 1788, although I suspect that this latter text is deliberately archaic. Both the Greek and Arabic astronomical numerals often co-occur in the same texts as pure decimal systems (either ciphered-additive ones such as alphabetic numerals, or ciphered-positional ones such as Arabic or Western positional numerals).

The sexagesimal fractions were only ever used by astronomers (and mathematicians working with astronomical problems), and even so, when writing non-astronomical material, they used ciphered-additive alphabetic systems. Yet, for this limited set of functions, they were used continuously for well over 1000 years. I believe that the best explanation for their survival is that they were very useful for the particular types of calculations required by astronomers. That sexagesimal notation was used solely for astronomy suggests that the demands of the discipline led to its retention. The division of the circle into 360 degrees (with subdivisions of 60 minutes per degree and 60 seconds per minute) is very useful, since 60 has a very large number of divisors.[6] Faddegon (1932) showed that this feature enables quick and easy multiplication and division using sexagesimal fractions, and cites it as one of the reasons for its retention among Arab astronomers after the positional Arabic numerals were introduced. Sexagesimal fractions were not simply a representational system – their utility was connected to computations involving a specific metrological system (in fact, the related systems for measuring angles of a circle and for time).

Although sexagesimal fractions are no longer used, modern astronomers still use the sexagesimal division of the circle, and anyone who can read a digital clock uses a kind of sexagesimal numeration. While we no longer mix additive and positional principles in notating time and angles, astronomers continue to restrict themselves to values under 60 for the division of the sky into segments, just as everyone is able to realize that thirty minutes pass between 1:50 and 2:20. These vestiges come to us, via Greek and Arabic sexagesimal fractions, from the Babylonian custom of numbering by 60. In this way, a peculiar custom of numeration, useful for astronomy but not much else, has had a significant impact on how humans perceive and structure time throughout most of the world today. However, these do not represent a sexagesimal numerical notation system, but simply a sexagesimal division of various metrological units (for angles and for time) that are then represented with decimal numerals. To write "11:05" does not mean 665 minutes (11x60+5), but simply 11 hours and 5 minutes. That we continue to measure time and angles in this way is a problem in the history of astronomy and of timekeeping that is well beyond the scope of this work.

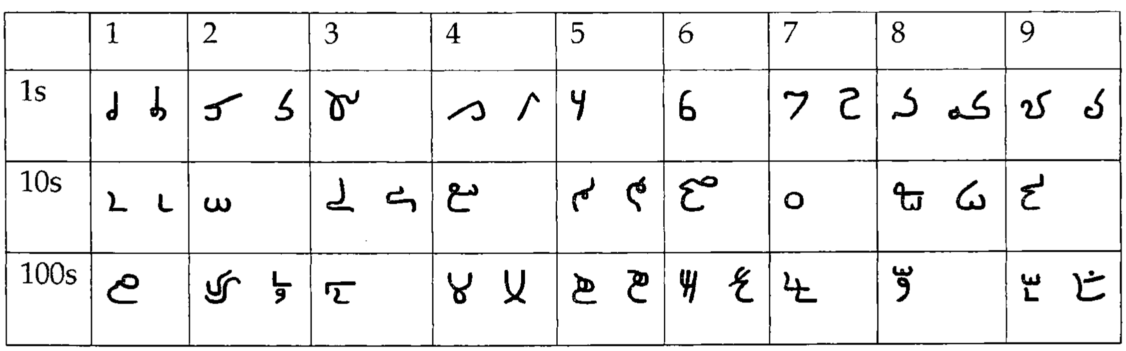
---

[6] In number theory, 60 is a "highly composite" number, defined as a natural number that has more divisors than all the numbers below it (Wells 1986: 127-128). 60 has twelve divisors: {1, 2, 3, 4, 5, 6, 10, 12, 15, 20, 30, 60}.

Fez numerals

A very interesting side note to the alphabetic family of numerical notation systems is found in the western extremity of the Muslim world, first briefly in medieval Spain, then around the city of Fez in modern Morocco starting in about the 16th century and continuing until very recently. While found only in Arabic manuscripts, the system is quite distinct both from the earlier Arabic abjad and from the Greek and other alphabetic systems in use at the time. This system was basically unknown to the West until 1917, when it was given the name "Fez numerals" or "Fez signs". The numeral-signs, including paleographic variants where appropriate, are indicated in Table 5.15 (Colin 1933: 199-201).

Table 5.15: Fez numerals



The system is decimal and ciphered-additive, and, like the Arabic script, is written from right to left with the highest values on the right. The twenty-seven signs are thus sufficient to express any number less than 1000. The resemblance between the signs for 6 and 7 and the modern Western numerals is likely a coincidence. For higher numbers, a subscript stroke placed to the left of any of the 27 signs indicates that its value should be multiplied by 1000. The Fez numerals thus constitute a hybrid multiplicative-additive system for values above 1000.

The origin of the Fez numerals is somewhat enigmatic. Colin notes that numerals very much like the Fez numerals were used among the Mozarabs (Arabic Christians) in Toledo, Spain in the 12th and 13th centuries (1933: 204). Levi Della Vida's study of these documents, which includes a table of these numeral-signs, confirms that they are essentially identical to the Fez numerals, except that they are written from left to right (Levi Della Vida 1934). The question then arises how these numerals came to be used in 12th century Spain. Three possible ancestors - the Arabic abjad (Maghreb variant), the Greek minuscule alphabetic numerals, and the Coptic "numerals of the Epakt" - are depicted alongside the Fez numerals in Table 5.16.

Table 5.16: Arabic abjad, Greek, Epakt, and Fez numerals

	Arabic	Greek	Coptic (Epakt)	Fez numerals
1	ا	$\bar{\alpha}$	[illegible]	[illegible]
2	ب	$\bar{\beta}$	[illegible]	[illegible]
3	ج	$\bar{\gamma}$	[illegible]	[illegible]
4	د	$\bar{\delta}$	[illegible]	[illegible]
5	ه	$\bar{\varepsilon}$	[illegible]	[illegible]
6	و	$\bar{\varsigma}$	[illegible]	[illegible]
7	ز	$\bar{\zeta}$	[illegible]	[illegible]
8	ح	$\bar{\varepsilon}$	[illegible]	[illegible]
9	ط	$\bar{\theta}$	[illegible]	[illegible]
10	ي	$\bar{\iota}$	[illegible]	[illegible]
20	ك	$\bar{\kappa}$	[illegible]	[illegible]
30	ل	$\bar{\lambda}$	[illegible]	[illegible]
40	م	$\bar{\mu}$	[illegible]	[illegible]
50	ن	$\bar{\nu}$	[illegible]	[illegible]
60	ص	$\bar{\xi}$	[illegible]	[illegible]
70	ع	$\bar{o}$	[illegible]	[illegible]
80	ف	$\bar{\pi}$	[illegible]	[illegible]
90	ض	ϙ	[illegible]	[illegible]
100	ق	$\bar{\rho}$	[illegible]	[illegible]
200	ر	$\bar{\sigma}$	[illegible]	[illegible]
300	س	$\bar{\tau}$	[illegible]	[illegible]
400	ت	$\bar{\upsilon}$	[illegible]	[illegible]
500	ث	$\bar{\phi}$	[illegible]	[illegible]
600	خ	$\bar{\chi}$	[illegible]	[illegible]
700	ذ	$\bar{\psi}$	[illegible]	[illegible]
800	ظ	$\bar{\omega}$	[illegible]	[illegible]
900	غ	ϡ	[illegible]	[illegible]

The situation is obviously a complex one. All four systems are written cursively and have an enormous amount of variation. It is possible that the Mozarabs' numerals are paleographic variants of the Greek alphabetic numerals, and thus came to the Arabs of Spain via direct diffusion from the Byzantines (Levi Della Vida 1934: 283). However, the situation is more complex, because many of the numeral-signs bear no resemblance to the Greek alphabet. Colin suggests, rather, that the Fez numerals (and their Spanish antecedent) were borrowed, not directly from the Greek alphabetic numerals used in the Byzantine Empire, but by way of the "numerals of the Epakt" used by Coptic Christians under Arab domination in Egypt (Colin 1933: 213). The remarkable paleographic similarities between several of the Coptic and Fez numeral-signs (e.g. 8, 80, and 500) suggest that some connection must exist between the two. Yet the hypotheses of direct diffusion from either the Greek numerals or the Epakt numerals suffer from the structural difficulty that, while the Fez numerals are multiplicative at only one level (1000), both of these candidates for its origin are multiplicative at both 1000 and 10,000.

I am unconvinced that the Fez numerals are solely derived either from the Greek alphabetic numerals or from the Copto-Arabic "numerals of the Epakt". It is reasonable to assume that both systems would have been known by the well-educated, privileged Mozarab Christians of Spain in the late 12th century. Yet a missing factor in this discussion is the role of the Arabic abjad used in the Maghreb in the development of the system. While the Arabic abjad numeration had fallen out of general use by the time the Fez numerals were developed, it was still used for number-magic and astronomical notation. Its use, though restricted to these rather abstruse topics, probably would have been familiar to the very individuals who would have developed the Fez numerals. Colin was aware of the Arabic abjad, but was unclear as to the time and region of its use, and accordingly did not consider it a possible ancestor to the Fez numerals (Colin 1933: 203). However, there are good reasons to postulate a connection. Some of the paleographic resemblances between the Fez numerals and the "numerals of the Epakt" (e.g., the signs for 7, 30, and 90) can be explained by the common factor of their relation to the appropriate letters of the Arabic abjad. Furthermore, unlike the other two systems, the Arabic abjad is multiplicative only in combination with the sign for 1000. Most likely, the Fez numerals are an unusual blend of the Greek, Coptic, and Arabic alphabetic systems adopted among a very unusual group of users, highly educated Arabized Christians living in Muslim-dominated southern Spain.

Yet these numerals did not last long in Spain; I know of no texts from the 14th century or later in which they are used. Instead, they spread into North Africa, probably after the expulsion of the Moors in 1492. While Colin notes that the first instance of the use of the numerals in North Africa was by the great historian Ibn Khaldun in the late 14th century, he argues from the textual evidence that it was not until the first half of the 16th century that their use in Morocco began in earnest (Colin 1933: 206). They appear to have been first employed in accounting and other commercial functions – a context which is very similar to that of the "numerals of the Epakt". They were used frequently throughout the 16th and 17th centuries, after which time they began to be replaced by the Arabic positional numerals.

At the time of Colin's paper in 1933, the Fez numerals were still in use for the single and quite limited purpose of indicating monetary values in wills. It is probable that the reason for their retention is twofold. Because inheritance values are often round numbers (multiples of 10), a ciphered-additive notation will generally abbreviate numeral-phrases as compared to a ciphered-positional system such as the Arabic or Western numerals. More importantly, because the meaning of the numerals was known to only a few specialists, the system serves as a cryptographic notation to prevent fraudulent modifications or forgeries (Colin 1933: 195). That this knowledge is the preserve of a learned few notaries suggests that the restriction of access to information is the primary function served by the Fez numerals in the modern period. No research on this system has been undertaken recently, but it seems unlikely that the system is still in use given the political changes that have taken place in post-colonial Morocco.

Armenian

Before the introduction of Christianity to Armenia, there was no native script in the region, and the Babylonian, Greek, and Old Persian scripts were used for literary purposes. The Armenian adoption of Christianity in the early 4th century AD was followed by enormous influence from the Greek-speaking world. In the early 5th century AD (probably in 406 or 407), the Armenian scholar-monk Mesrop Mashtots (c. 360 – 440) developed the first uniquely Armenian script, an alphabet of 36 letters, in order to translate the bible from Greek into Armenian (Sanjian 1996: 356).[7] At the same time, the letters of the alphabet were assigned numerical values as shown in Table 5.17.

Table 5.17: Armenian numerals

	1	2	3	4	5	6	7	8	9
1s	Ա	Բ	Գ	Դ	Ե	Զ	Է	Ը	Թ
10s	Ժ	Ի	Լ	Խ	Ծ	Կ	Հ	Ձ	Ղ
100s	Ճ	Մ	Յ	Ն	Շ	Ո	Չ	Պ	Ջ
1000s	Ռ	Ս	Վ	Տ	Ր	Ց	Ւ	Փ	Ք

The signs listed in the table above are similar to the uncial letters that were used exclusively from the 5th through 10th centuries AD, known as *erkat'agir* "iron-forged letters" (Thomson 1989, Sanjian 1996: 357). In the 11th century, cursive minuscule letters known as *bolorgir* began to be used; these are now the standard forms used in printed Armenian books. The system is ciphered-additive and decimal, and is written from left to right. Because the ancient Armenian alphabet had 36 letters, it had enough signs to express the complete series from 1000 to 9000 as well as all the units, tens, and hundreds.

[7] The modern Armenian script has 38 letters, the last two of which (*o* and *fē*) were introduced in the medieval period and have no numerical value.

The system could thus denote any values less than 10,000. However, unlike many ciphered-additive systems of this family, the Armenian system does not use multiplication at 1000, 10,000, or any other point in order to express higher values, which were written out in full using lexical numerals.

Our information on the history of the ancient Armenian numerals is restricted by the fact that very little epigraphic or paleographic evidence survives from the earliest centuries of the system's use. Nevertheless, it is certain that the Armenian numerals were developed on the model of the Greek alphabetic numerals, just as the Armenian script itself was derived from the Greek. Many other scripts, including Syriac, Aramaic, Ethiopic, Pahlavi, and Phoenician, have been suggested as possible ancestors of the Armenian script, based on resemblances in the shapes of certain characters (Gamkrelidze 1994: 37). In part, disagreements over the origin of the script rest on the fact that there are few or no resemblances in the form of the Armenian letter-signs with those of Greek. However, of the possible ancestors of the Armenian script, only the Greek alphabet used appropriate alphabetic numerals. While the Ethiopic script used ciphered-additive numerals, they are very different structurally from the Armenian numerals, and do not represent phonetic characters. The Syriac script did not employ alphabetic numerals until around 600 AD, before which time the Syrians used a cumulative-additive system of the Levantine family (ch. 3). While other scripts may have been used to borrow a few signs, the primary stimulus for the invention of the Armenian script and numerals was direct contact with Greek speakers, probably in the context of missionization. Many of the script signs probably were designed independently of any particular external influence, in the interest of concealing rather than highlighting the connection between the Greek and Armenian alphabets.

It is unclear whether the Armenian alphabetic numerals were developed by Mesrop Mashtots himself (or his assistants) in the early 5th century AD, or whether they were produced later in the century. Still, given that their development came within a few decades of the script's origin, it is reasonable to assert that Mashtots or his immediate successors were responsible for its invention. The system itself has obvious structural similarities with all the ciphered-additive systems of this family. Yet because the Armenian system uses the last nine signs of the alphabet as signs for 1000 - 9000, its inventor(s) must have recognized the sufficiency of the Armenian alphabet's 36 signs for expressing the thousands values, and thus modified the hybrid multiplicative Greek system into a purely ciphered-additive one. The primary weakness of this new structure was that it could not be used to express numbers above 10,000, but this was not a concern in the early centuries of the script.

Although a connection is sometimes asserted to exist between the Armenian and Georgian alphabetic numerals, the evidence for this is too tenuous to suggest any definite link. The primary similarities between the two are that both were used in the same region and had distinct signs for 1000-9000. The only system that is clearly derived from the Armenian alphabetic numerals is the variant Armenian system developed in the 7th century AD by Anania Shirakatsi, which I will discuss below. The Armenian numerals did not spread beyond the limited area around Lake Van where the Armenian language was spoken, nor do they appear to have inspired the creation of any foreign systems. After the development of the minuscule Armenian numerals, these signs were also used numerically in the same way as the older uncial signs.

Ciphered-positional numerals - the Arabic system used by the neighbouring Seljuk Turks - were first used in Armenia in the 12th century (Shaw 1938-9: 368). Yet Armenian writers retained the alphabetic numerals for most ordinary purposes long afterward. Only in the mid-17th century, when Armenia had been firmly under Ottoman control for some time, did ciphered-positional numerals replace the alphabetic system. The Armenian alphabetic system was still used in 1911 for numbering chapters of the New Testament, although page and verse numbers were written using Western numerals. The numerals used in modern Armenia are the standard Western numerals.

Shirakatsi's notation

The Armenian astronomer, geographer, and mathematician Anania Shirakatsi[8], who lived in the 7th century AD, was Armenia's greatest pre-modern scholar. While little known today outside his native country, Shirakatsi's contribution to Armenian learning is unparalleled, particularly through his synthesis of Persian, Arabic, Greek, and other scientific knowledge. In addition to these accomplishments, Shirakatsi developed a very interesting numerical notation system that is structurally different from any other system used in the region. In Western scholarship, this system has only been examined in a brief paper by Shaw (1938-39), and thus has never been considered in the standard literature on the history of numerals. The basic form of this system uses only twelve signs, as shown in Table 5.18 (Shaw 1938-9: 270).

Table 5.18: Armenian numerals: Shirakatsi's notation

	1	2	3	4	5	6	7	8	9
1s	Ա	Բ	Գ	Դ	Ե	Զ	Է	Ը	Թ
10	Ժ		100	Ճ		1000	Ռ		

These signs are identical to those used for the appropriate numbers in the traditional Armenian system. However, Shirakatsi provided many examples explaining how these signs could be combined to express numbers through multiplication as well as addition. In this system, a unit-sign followed by one of the three exponent-signs (for 10, 100, or 1000) indicates that the values of the two should be multiplied; these pairs of signs could be put together into a larger numeral-phrase through addition. Instead of writing 9642 as ՔՈԽԲ (9000+600+40+2), as in the traditional Armenian alphabetic numerals, Shirakatsi would write the same number as ԹՌԶՃԴԺԲ (9 x 1000 + 6 x 100 + 4 x

[8] Also known as Ananiah Shiragooni, or Ananiah of Shirag.

10 + 2). Thus, where the traditional Armenian system is ciphered-additive, Shirakatsi's system is multiplicative-additive.

It is immediately obvious, however, that any numeral-phrase can be written more compactly with the traditional alphabetic numerals than with this multiplicative-additive variant. Why, then, would Shirakatsi advocate this system's use? Firstly, it requires knowing fewer symbols (12 versus 36) in order to express any number less than 10,000. The importance of this factor is minimized by the fact that the system's users would already know the 36 classical Armenian letters and their order. Shirakatsi also showed how numbers greater than 10,000 could be expressed using multiplicative combinations of two or three signs. To do so, however, one needs the entire repertoire of Armenian numerals from 1 through 9000, as described earlier. For numbers from 10,000 through 90,000, Shirakatsi juxtaposed the signs for 10-90 with the sign for 1000. Similarly, the numeral-phrases for 100,000 through 900,000 combined the signs for 100-900 with the sign for 1000. Alternately, the hundred thousands could be expressed using unit-signs followed by a 100-sign and then a 1000-sign. Thus, one could write 460,000 as ՆԿՌ – (400+60) x 1000 – or ԴՃԿՌ – ((4x100) + 60) x 1000. This system is no longer a purely decimal system, but has a mixed base of 10 and 1000. For values below 1000, it is purely multiplicative-additive, but above 1000, the multiplicand that is juxtaposed with the sign for 1000 (Ռ) is not a single sign, but rather a ciphered-additive *numeral-phrase*.

Because so little is known about this system, speculations on its origin will be tentative at best. Obviously, it existed in the 7th century AD at the time Shirakatsi was writing, but Shaw believes that it was not developed by Shirakatsi, but was a commonly used Armenian variant system, of which Shirakatsi's writings are the only surviving remnant (Shaw 1938-9: 369). I do not believe there is any reason to regard the system as anything other than the creation of Shirakatsi himself, since its structure is never found in Greek, Syriac, Hebrew, or any other system of the alphabetic family. The intriguing possibility exists that Shirakatsi borrowed the notion of multiplicative structuring from one of two foreign sources. The interesting numerals developed by the 5th century AD Indian mathematician Âryabhata (ch. 6) were multiplicative-additive; it is possible that Shirakatsi, a mathematician with extensive knowledge of foreign writers, knew of Âryabhata's numerals and emulated them. Similarly, the traditional Chinese numerical notation system (ch. 8) is multiplicative-additive, and it is plausible that Shirakatsi knew of it. Nevertheless, particularly because of the unusual way numbers above 10,000 are expressed, I believe that the unusual structural modifications to the Armenian numerals were developed by Shirakatsi himself.

Shirakatsi's system is thus a structurally innovative variant of the Armenian numerals designed to facilitate the representation of large numbers of the sort that would be needed for his astronomical and mathematical calculations. Yet I do not agree with Shaw's unusual theory that Shirakatsi's system is the ancestor of the Western numerals (Shaw 1938-9: 371-2). Shaw does not seem to be familiar with the distinction between a multiplicative-additive system such as Shirakatsi's, which does not use the positional principle, and ciphered-positional systems such as the Hindu, Arabic, and Western positional systems. Shirakatsi's system is structurally far closer to the traditional (multiplicative-additive) Chinese system than it is to any positional system. There is no evidence that his system was adopted by any later writers, or that it had any effect on the development of other numerical notation systems throughout the world. Instead, we should view this system as the creative invention of a single individual, used only within his lifetime.

Georgian

Like the Armenians, the Georgians developed a script and numerical notation system modelled after the Greek alphabet shortly after the adoption of Christianity in the region. The creation of this first Georgian alphabet is often attributed in folklore to King Parnavaz, and is sometimes said to have been created in the 3rd century BC; however, this theory does not withstand serious scrutiny. While Christianity was adopted in Georgia in 337 AD, there is no direct evidence of Georgian writing until about a century later, at which time the *asomtavruli* or majuscule script began to be used (Holisky 1996). More familiar to modern scholars, however, are the *mxedruli* characters developed in the 11th century AD, which continue to be used to write the Georgian language today. The numerals associated with this script are shown in Table 5.19 (Holisky 1996: 366).

Table 5.19: Georgian numerals

	1	2	3	4	5	6	7	8	9
1s	ა	ბ	გ	დ	ე	ვ	ზ	ჱ	თ
10s	ი	კ	ლ	მ	ნ	ჲ	ო	პ	ჟ
100s	რ	ს	ტ	უ	ფ	ქ	ღ	ყ	შ
1000s	ჩ	ც	ძ	წ	ჭ	ხ	ჴ	ჯ	ჰ

The system is decimal and ciphered-additive, and, like the Georgian script, is written from left to right. Thus, 4808 would be written as წყჱ. Like the Armenian script, the Georgian script had enough letters to serve for all numerical values up to 9,000. Some later inscriptions even include a special sign for 10,000 (ჵ). There is no evidence that the Georgian alphabetic numerals were ever used to express higher numbers than this, either through multiplication or through additional signs. Presumably, such numbers were written out in full using lexical numerals.

We are faced with the nearly insurmountable problem that the early history of the Georgian numerals, like that of the Armenian numerals, is cloudy at best. Were the Georgian numerals developed immediately upon Christianization in the mid-4th century AD, or not until the 5th century or even later, when the earliest Georgian inscriptions are found? Are there external influences on the Georgian script and numerals other than from Greece, such as from Armenian or Hebrew? Was the system's inventor a native

Georgian or Greek? The answers to these tantalizing questions unfortunately remain incomplete, but a few facts have become clear.

It is often asserted by Armenian sources that, after creating the Armenian script, Mesrop Mashtots modified his invention for use in writing the Georgian language. This conclusion is supported by the geographical proximity of Armenia and Georgia, and the fact that the exact circumstance of the development of both scripts is shrouded in mystery. With respect to the numerical notation systems, there is an undeniable structural similarity between the Georgian and Armenian systems, which both, unlike the Greek alphabetic numerals, have enough additional letters to represent the values from 1000 through 9000. However, the major problem with this theory, highlighted by Gamkrelidze, is that while the Georgian and Armenian scripts both use 36 signs for 1 through 9000, the letter-order of the two scripts is vastly different. The Georgian letter-order was modelled very closely on the Greek, with additional signs added as necessary at the end of the series, while the uniquely Armenian phonemes in that script were interspersed almost randomly within the original Greek letter-order. Gamkrelidze points out that while both the Georgian and Armenian numerals (and scripts) are almost certainly based on the Greek system, it is unlikely that the Georgian numerals would be modelled on the Armenian numerals but retain the Greek letter-order for their values (1994: 77). He concludes, *a fortiori*, that Mesrop Mashtots was certainly not the inventor of the Georgian script and numerals. He admits that there may have been some mutual influence between the two scripts and numerical notation systems, given certain similarities in the sign-forms, but rightly notes that the direction of this influence remains unclear (Gamkrelidze 1994: 81-82). The possibility that the Georgian numerals are *ancestral* to the Armenian numerals has not been addressed in present scholarship, and remains an intriguing area for future research.

It is likely that the Georgian alphabetic numerals were developed on the model of the Greek alphabetic numerals, but independently of the Armenian numerals. If

Gamkrelidze is correct, the development of an indigenous Georgian script and numerical notation system whose signs are largely independent from external influences was motivated by the desire to create something useful and uniquely Georgian (1994: 68). While, in contrast to the Armenian situation, the order of the Greek numerals was retained for the most part, additional letters were used to complete the system up to 9,000.

The Georgian numerals were used in literary and religious texts throughout the medieval period, particularly for pagination, dating, and stichometry, and were used on a large number of monumental inscriptions. Their employment as a regular system ended in the 16th century, at which time Georgia came under Ottoman control and the Arabic positional numerals were used for administrative and commercial purposes, although the alphabetic numerals may have been retained for religious functions. However, Paolini and Irbach's 1629 Georgian-Italian dictionary, the first book printed in Georgian, does not contain any mention of the alphabetic numerals alongside its list of Georgian letters. Since the 18th century, when Georgia fell under the Russian sphere of influence, the Western numerals have been those normally used for all purposes in written Georgian.

Glagolitic

The Glagolitic script was probably developed between 860 and 870 by the brothers Cyril and Methodius who, while on a mission to the Moravian Slavs of what is today modern Serbia, Croatia and Macedonia, created an alphabet for Slavic liturgical writings in the language now known as Old Church Slavonic (Schenker 1996: 166-7). There may have been a pre-Christian script in the region, which might explain why many of the Glagolitic letters have no correlation with the Greek alphabet (Cubberley 1996). This is certainly not the case with the Cyrillic alphabet also invented by Cyril and

Methodius, whose letters are clearly derived from Greek. Regardless of the ultimate origin of the Glagolitic script, it is clear that the Glagolitic numerical notation system belongs to the alphabetic family. The numeral-signs of the system are shown in Table 5.20 (Vaillant 1948; Gardiner 1984).

Table 5.20: Glagolitic numerals

	1	2	3	4	5	6	7	8	9
1s	Ⰰ	Ⰱ	Ⰲ	Ⰳ	Ⰴ	Ⰵ	Ⰶ	Ⰷ	Ⰸ
10s	Ⰻ	Ⰹ	Ⰼ	Ⰽ	Ⰾ	Ⰿ	Ⱀ	Ⱁ	Ⱂ
100s	Ⱃ	Ⱄ	Ⱅ	Ⱆ	Ⱇ	Ⱈ	Ⱉ	Ⱋ	Ⱌ

As with the Greek and many other systems, Glagolitic numerals were frequently distinguished from words in texts by placing dots to either side of a numeral-phrase or by placing a mark of some sort above it (Vaillant 1948: 24; Schenker 1996: 182). In addition to these 27 signs, additional signs for 1000, Ⱍ, and 2000, Ⱎ, were used in some texts. The system is ciphered-additive and decimal, and is always written from left to right. However, for the numbers 11 through 19, the ordinary sign order is reversed (e.g. ⰁⰋ instead of ⰋⰁ for 12), which reflects the Slavic numeral words for the teens (Schenker 1996: 182).

There is considerable confusion concerning the higher Glagolitic numerals. Apparently, none of the surviving Glagolitic manuscripts show any indication of using numerals higher than 1000 (Gardiner 1984: 15; Lunt 2001: 28). Yet Schenker contends that the Glagolitic thousands were expressed by placing a small diagonal or curved stroke (like the Greek *hasta*) to the left of a numeral-sign to indicate that its value should be multiplied by 1000 (1996: 182). I do not know on what basis he contends this; perhaps it is on the model of the Cyrillic numerals, which do use such a sign. If Schenker is correct, Glagolitic is a hybrid multiplicative-additive system above 1000. Finally, Gamkrelidze and others contend that, because the earliest Glagolitic script had 36 characters, it is likely that the last nine letters of the alphabet (of which most were later dropped from the script) originally had the values 1000-9000 (Gamkrelidze 1994: 39-40). At present, I think it safest to regard the expression of the thousands in Glagolitic as insufficiently common to advance any definite conclusion as to how they were written, although I think the conclusion of multiplicative structuring with a *hasta* is most likely.

It is likely that the Greek alphabetic numerals were the sole influence on the origin of the Glagolitic numerals. The similarities in structure between the Greek and Glagolitic systems, coupled with the fact that Cyril and Methodius were Greeks, renders impossible the hypothesis of its independent invention. As for other alphabetic systems as possible candidates for its origin, the Gothic numerals were long defunct by the 9th century AD and the Cyrillic numerals were not invented until later in the century. The question of the ultimate origin of the Glagolitic alphabet and, thus, the numeral-signs, is an interesting one. Schenker concludes that a variety of scripts, such as the Latin, Greek, Samaritan, and Hebrew, may have been used as the model for one or more signs, with other signs being unique inventions with no obvious correlates in other scripts (Schenker 1996: 168-172). Even if this is the case, the Glagolitic letters must have been assigned their numerical values under the influence of Greek Christianity.

The later history of the Glagolitic numerals is marked by its slow replacement by neighbouring systems, such as the Cyrillic and Roman numerals. Manuscripts were written in Glagolitic throughout the medieval period in the region of modern Croatia, Serbia, Slovakia, and even into the Czech Republic and Poland. Yet, even during the Middle Ages, Catholic or Western-influenced areas began to prefer the Roman numerals to the Glagolitic, while areas under Bulgarian or Serbian control tended to adopt the Cyrillic numerals and script. Russian writers never used the Glagolitic signs, and always used the Cyrillic alphabetic system. By the 15th century, almost all the Slavs had adopted either Roman or Cyrillic numerals.

Only in Croatia, particularly along the Adriatic coast (Dalmatia), did the Glagolitic script and numerals flourish. Glagolitic was specifically retained for the Croatian Roman Catholic liturgy (Cubberley 1996: 350). It was also used in a variety of monumental contexts in Croatia from the 11th century onward, a context not seen elsewhere. Croatian is the only language for which the Glagolitic script was used for printed books. Yet, even in Croatia, the Glagolitic script and numerals declined greatly in use after the Ottoman conquests of the 16th century, and were used only rarely from the 17th century onwards (mostly in religious texts). It is not clear whether the Glagolitic numerals survived as long as the Glagolitic script, which persisted until the beginning of the 20th century in the islands of the Quarner archipelago in northwestern Croatia.

Cyrillic

Like Glagolitic, the Cyrillic script was developed under the guidance of the missionaries Cyril and Methodius. It is quite likely that the development of Cyrillic was accomplished, in fact, after the death of Cyril and Methodius, by Cyril's followers and disciples in Bulgaria in the 890s AD, who then named the script after their deceased mentor. Cyrillic was used for writing the Old Church Slavonic language used for Slavic liturgical texts, and later was adopted for writing a variety of Slavic languages. Alongside the Cyrillic script, an alphabetic numerical notation system was developed around the same time. Its numeral-signs are shown in Table 5.21 (Gardiner 1984: 16-17; Cubberley 1996: 348).

Table 5.21: Cyrillic numerals

	1	2	3	4	5	6	7	8	9
1s	а	В	Г	Д	Є	Ѕ	Ꙁ	И	Ѳ
10s	І	К	Л	М	Н	Ѯ	О	П	Ч
100s	Р	С	Т	ОУ	Ф	Х	Ѱ	Ѡ	Ц

The system is ciphered-additive and decimal, and is normally written from left to right. For the numbers 11 through 19, the ordinary sign order was often reversed (e.g. ВІ instead of ІВ for 12), which reflects the structure of Slavic words for those numbers (Vaillant 1948: 24). Numeral-phrases were often distinguished from ordinary letters by placing a bar or other mark above the phrase, and sometimes also by placing dots on either side of the signs (Lunt 2001: 28). Placing a small stroke to the left of a number indicated that its value should be multiplied by 1000 (Schenker 1996: 182; Vaillant 1948: 24). The Cyrillic numerical notation system is thus a hybrid: purely ciphered-additive below 1000 and multiplicative-additive for higher exponents. Gardiner states that the multiplicative Cyrillic numerals were expressed by using the units preceded by an unusual sign, ҂, to indicate multiplication by 1000 (Gardiner 1984: 15). To my knowledge, such a sign was never used with any phonetic value in Cyrillic and I cannot confirm its use in any document.

While there are only 27 signs listed above, there are more than 27 signs in all varieties of the Cyrillic script; modern Russian Cyrillic uses 32 letters and earlier Cyrillic scripts used a number of older signs that have now fallen into disuse. The signs that are assigned numerical values in Cyrillic are those which are directly derived from Greek, including the otherwise rarely used signs for xi (Ѯ), psi (Ѱ), and theta (Ω). Yet numerical values were never assigned to the commonly used but non-Greek characters (Gardiner 1984: 14-15). Thus, the Cyrillic numerical values do not correspond to the customary order of letters in several respects. In general, the Cyrillic numerals and script are far more faithful to the original Greek than Glagolitic.

The circumstances of the origin of the Cyrillic script and numerals are better understood than for almost any other script. Its first appearance dates to around 890 AD, at which time Slavs and Greeks who had been influenced by Cyril and Methodius were extremely active in the Christianization of the Slavs in the region of modern Bulgaria. That this missionary work was undertaken under the auspices of the Byzantine Empire confirms what is clear from the paleographic evidence - that the sole external influence on the Cyrillic script and numerals was the Greek uncial alphabet used at the time. The non-Greek signs developed to express additional consonantal Slavic phonemes were never assigned numerical values, further confirming the Greek origin of the Cyrillic numerical notation system.

From its origins in Bulgaria and Serbia, the Cyrillic numerical notation system spread to Kievan Rus in the 10th century. The Balkans fell under Ottoman influence in the 15th century, after the fall of Constantinople, and the alphabetic numerals had generally ceased to be used there by around 1500. In Russia, the alphabetic numerals were used much longer; not until the reforms of Peter the Great around 1700 were the Western positional numerals introduced on a widespread basis as the numerals of administration, law, and commerce. However, unlike in the West, where the Western numerals were resisted for centuries, the transition from the alphabetic to the Western numerals appears to have taken place quite rapidly in Russia as the desire of a single individual (Peter I) to change the notation system, without external political domination. Today, the Cyrillic numerals are occasionally used in modern Church Slavonic texts (especially for numbering chapters and verses in Bibles), but never occur in ordinary Cyrillic writing (Gasparov 2001: 17-18).

Summary

The alphabetic family originated with the Greeks in the 6th century BC, who combined the structure of the Egyptian demotic system with the idea of using phonetic signs in an assigned order as numeral-signs. The political and religious power of the Greek-speaking world (particularly during the Byzantine period), coupled with the brevity and adaptability of ciphered-additive alphabetic numerical notation systems, led to the development of other alphabetic systems modelled on the Greek numerals but using numeral-signs specific to each script. This family expanded tremendously between the 4th and 7th centuries AD (the time of greatest Eastern Roman / Byzantine power), with eight new systems arising during this period. Yet most systems of this family had died out, or at least were greatly reduced in the contexts of their use, by the 16th century AD, during which time the Arabic positional and Western numerals took a firm place as the ordinary numerals of commerce and administration throughout Europe and the Middle East. Many systems of this family are still used today, but are used only in limited contexts (e.g. liturgical texts or number-magic). Yet the interest of the alphabetic family of numerical notation systems is not simply in its culture history; rather, a number of intriguing theoretical insights can be gained from a comparison of the systems of this family.

There is no one feature common to all the systems of this family. The most common structure is ciphered-additive with a decimal base, with or without the use of multiplicative-additive structuring for the higher exponents. However, the Armenian notation of Shirakatsi is multiplicative-additive and sometimes uses a base-1000, while the Greek and Arabic astronomical notations are quasi-positional and involve a sexagesimal base. With the exception of these two systems, which were designed for specific mathematical and astronomical purposes, all other systems conform to the basic decimal ciphered-additive framework. Even so, there is enormous diversity among the remaining systems – in the number of signs used, the way in which multiplication is or is not used to express higher exponents, and whether or not the signs used correspond with the script-signs of the language in question.

It is unsurprising that the inventors of a numerical notation system of this family would use local script-signs rather than those of the system's ancestor. One of the great advantages of alphabetic systems is that, if the signs are ordered according to the values of a local script, one need not learn both a set of script signs and a set of numeral-signs; one merely superimposes the decimal structure of the numerals onto the script.

Combining these two functions into one single system lessens the mnemonic burden on new learners of a numerical notation system as well as on experienced users. The Greek alphabetic, Coptic, Gothic, Hebrew, Syriac, Armenian, Georgian, and Glagolitic systems all take advantage of this feature. Nevertheless, this feature is not universal within the alphabetic family. The values assigned to Arabic and Cyrillic letters do not correspond to the customary letter-order, thus reducing this benefit. The Fez numerals and the Coptic Epakt numerals are blended alphabetic systems, combining the numeral-signs of two or more existing alphabetic systems to create a new system that does not represent the signs of any script. In the Ethiopic system, the users of one script adopted the ordered numeral-signs of another (in this case, the Greek alphabet) rather than adopting both the script and numerals. By failing to do so, the inventors of the Ethiopic system deprived themselves of the useful alphabetic convenience of most systems of this family.

Despite the advantage of combining phonetic and numerical representation systems, the numerical notation systems of this family require many signs. Comparing the alphabetic systems with the cumulative-additive systems of Chapters 2 through 4 shows the much greater number of numeral-signs used in alphabetic systems. Even the Ethiopic system, which is multiplicative above 100, requires 19 separate signs, more than any cumulative-additive system, and the Armenian and Georgian systems require as many as 36 signs. It is an inevitable consequence of their decimal ciphered-additive structure that these systems require nine signs for each exponent: 27 signs to express all numbers up to 1000. In the case of the Hebrew and Syriac systems, whose scripts only had 22 signs, numerals above 400 were expressed through cumulative combinations of hundred-signs. While this solves the problem of having only 22 signs in the system's repertoire, it makes numeral-phrases longer and more complex.

As it is inconvenient to develop nine new signs for each higher exponent of 10, many alphabetic systems are ciphered-additive for lower exponents but begin to use multiplicative-additive structuring above some specific point. The Gothic, Armenian, and

Georgian systems do not use multiplication at all, and thus are restricted to expressing numbers below 1000 (Gothic) or 10,000 (Armenian and Georgian), with all larger numbers expressed in writing. The Ethiopic system is multiplicative above 100, a feature that can only exist because the signs of the system are not phonetic signs of the Ethiopic script. A large plurality of systems: Cyrillic, Hebrew, Fez numerals, Coptic, Arabic abjad, and possibly Glagolitic - use multiplication above 1000, a natural way to proceed in systems with 27 ordinary signs. Three other systems: the Greek, the Syriac, and Coptic Epakt numerals, are multiplicative *both* at 1000 and 10,000; that is, after 8000 (8 x 1000) and 9000 (9 x 1000), one uses a new sign for 10,000 (1 x 10,000) rather than (10 x 1000) as in systems that are only multiplicative at 1000. The most reasonable hypothesis for why this feature would develop is that the Greeks (the first group to use it) borrowed their numerals from the Egyptian demotic numerals, which are multiplicative at 10,000 but *not* at 1,000. Because the Greek system only had 27 signs, they needed to add an extra level of multiplication at 1,000 in order to cover all numbers. Interestingly, only one system – and a very obscure one, Shirakatsi's numerals, takes the step of rendering the entire system in a multiplicative-additive fashion. The rarity of this approach is probably because multiplicative-additive numeral-phrases are usually longer than ciphered-additive ones. Thus, the structural homogeneity of a purely multiplicative-additive system, while perhaps appealing from an aesthetic point of view, has clear disadvantages in terms of compactness.

The longevity of the systems of this family is quite remarkable. Eight of the alphabetic numerical notation systems were regularly used for 1000 years or more (Greek, Coptic, Ethiopic, Hebrew, Syriac, Armenian, Georgian, and Cyrillic). In some cases, such as the Greek system, this can in part be explained by the lack of functional equivalents and/or the political importance of the system's users. In others - Hebrew, Armenian, and Georgian, for instance - the systems' users have largely been marginalized peoples. That these systems could survive - sometimes for centuries - in such sociopolitical circumstances and where, in many cases, functionally equivalent or superior ciphered-positional systems were available, requires some explanation. Firstly, as mentioned above, the "alphabeticity" of alphabetic numerical notation systems means that one need not learn a set of numerals in addition to a given script, which is not true of systems such as our own Western numerals. Secondly, ciphered-additive systems are always more compact than ciphered-positional systems. For any Western numeral-phrase containing zeroes, the corresponding ciphered-additive numeral-phrase will be shorter.

While the advantages of alphabetic systems are thus clear, I believe the most important reason why so many of these systems have survived so long is that alphabetic numerical notation systems, like scripts themselves, can be important markers of cultural identity. In many cases throughout this chapter (e.g. Coptic, Gothic, Armenian, Georgian, Glagolitic, and Cyrillic), a group of people developed a unique set of alphabetic numerals at or around the same time as they developed their own script. In other cases, such as the Hebrew or Syriac numerals, the relevant script had existed for centuries with a different numerical notation system before combining the older script-signs with the ciphered-additive structure of a foreign system. The point of alphabetic numerals is not, as with the Roman numerals, to be comprehensible trans-linguistically, but rather for each system to serve for one script alone. Under these circumstances, an alphabetic numeral system becomes an integral part of a script, and thus can be used to mark ethnic identity. Even when these systems cease to be used regularly, many of them continue to be used in restricted functions, particularly in the domain of religion (e.g. Hebrew, Syriac, Coptic, Greek). Because they continue to be used in these culturally sensitive contexts even after having been abandoned for purposes such as mathematics and commerce, it is reasonable to conclude that the cultural meaning they hold for their users may be the central reason for their retention.

Chapter 6: South Asian Systems

The South Asian family of numerical notation systems includes all systems that derive from the Brāhmī numerals used on the Indian subcontinent. While this family has its roots in India, its geographical distribution exceeds that of any other, given that its most widespread and prolific descendant is the Western system. With the possible exception of China, numerical notation systems of this family are predominant throughout the entire world today. As with many of the families I have discussed, the origin of the South Asian numerals is a contentious issue, albeit one that I hope can be clarified by considering both the structural features and historical context of their invention and development. The history of the family extends back at least to the 3rd century BC (possibly slightly earlier), and, of course, several of the systems of this family (Western, Arabic, and various Indian systems) remain in regular use to this day.

While most of the modern systems of this family are ciphered-positional - that is, their structure is that of our own numerals - there is much variability among the earlier numerals, including ciphered-additive and multiplicative-additive systems. An important evolutionary development in this family's history was the shift from ciphered-additive systems, such as the early Brāhmī numerals, to ciphered-positional systems. Despite repeated attempts to postulate the origin of the important ciphered-positional structure elsewhere (Greece, China, or Mesopotamia), that this development came out of South Asia can no longer be doubted.

Brāhmī

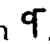
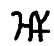
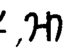
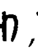
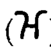
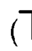
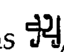
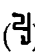
The Brāhmī script probably developed around 400 BC or slightly earlier, as is attested from inscribed potsherds from the site of Anuradhapura in Sri Lanka (Coningham *et al.* 1996). It came to prominence in the mid-3rd century BC, during the reign of the Mauryan emperor Aśoka. It was probably derived from a Semitic prototype

(Aramaic, South Semitic, or Phoenician), although many South Asian scholars still support the theory that the script was indigenously developed (Salomon 1996: 378-9). Brāhmī, along with the slightly earlier Kharoshthi script used in the northwestern regions of India, was the first script used in India since the collapse of the Harappan civilization. Both scripts are alphasyllabaries (scripts in which each sign has a consonantal base that is modified to indicate which vowel sound is associated with it). There are many structural differences between the two that suggest that their origins are quite possibly different, particularly that Brāhmī was written from left to right while Kharoshthi was written from right to left,. This is supported by the fact that the Kharoshthi numerals (ch. 3) are a hybrid cumulative-additive / multiplicative-additive system very much like Aramaic and related systems. In contrast, the Brāhmī numerals are quite different in principle. The basic Brāhmī numerals are shown in Table 6.1 (Salomon 1998: 58; cf. Datta and Singh 1962 [1935]: Table III-IX; Bühler 1896: Plate IX).

Table 6.1: Brāhmī numerals

	1	2	3	4	5	6	7	8	9
1s	𑁒	𑁓	𑁔	𑁕	𑁖	𑁗	𑁘	𑁙	𑁚
10s	𑁛	𑁜	𑁝	𑁞	𑁟	𑁠	𑁡	𑁢	𑁣
100s	𑁤	𑁤𑁓	𑁤𑁔	𑁤𑁕	𑁤𑁖				
1000s	𑁥	𑁥𑁓	𑁥𑁔	𑁥𑁕	𑁥𑁖				

The signs shown above are those found in Ksatrapa coins (2nd to 4th century AD), and are representative of the Brāhmī system throughout its history, although there are considerable paleographic variations among inscriptions. Numeral-phrases were written from left to right, proceeding from higher to lower exponents; thus, 289 might be written as 𑁤𑁓𑁢𑁚. The signs for 1 through 3 are essentially cumulative in nature, with horizontal strokes indicating units. Other than this minor cumulative component, the Brāhmī system is ciphered-additive up to 100, because it has separate signs for each of the units and the tens. The single-stroke signs found in early inscriptions become ligatured or distorted in many later inscriptions. For instance, the 7th/8th century AD grants of the Gangâ dynasty contain  and for 2 and 3 (Datta and Singh 1962 [1935]: Table V].

Above 100, the structural classification of the system becomes more complex. The numeral-signs in Table 6.1 are not, as they might appear, simple multiplicative formations that juxtapose a unit-sign with an exponent-sign for either the hundreds or the thousands; otherwise, we would expect  for 2000 rather than . These numeral-signs for 100-300 and 1000-3000 are not exactly multiplicative; we might call them *quasi-multiplicative*. Still, we cannot ignore the graphic similarity of the various signs for the hundreds and thousands to the corresponding units. At the early site of Nana Ghat (1st century BC), the numbers 400, 700, 1000, 4000, 6000, 10,000, and 20,000 are written in a clearly multiplicative fashion as , , , , , , and , and thus combine unit-signs with signs for 100 () and 1000 () (Indraji 1876). Similar multiplicative structures for the hundreds and thousands occur on inscriptions throughout the entire lifespan of the Brāhmī numerals. The Vâkâtaka grants (5th century AD), one of the latest texts containing signs for the thousands, expresses 8000 as , a ligature of the signs for 1000 () and 8 () that occur on the same grants (Datta and Singh 1962 [1935]: Table IV, Table IX).

This evidence suggests that, despite paleographic changes, the basic structure of the Brāhmī numerals was ciphered-additive below 100 and multiplicative-additive at both 100 and 1000. Since 10,000 and 20,000 are written as 1000 x 10 and 1000x 20 rather than 10,000 and 10,000 x 2, we know that no special sign for 10,000 was used. In the Nana Ghat inscriptions, 24,400 is written as  (1000 x 20 + 1000 x 4 + 100 x 4). It is possible that earlier, purely multiplicative forms existed, but are not known from any surviving texts. Additionally, it is possible that this structural feature informs us about the origin of the Brāhmī numerals.

The Brāhmī numerals appear on some of the earliest Aśoka inscriptions, dating to the middle of the 3rd century BC, but not in the early Sri Lankan writings. These early inscriptions contain only a few signs (for 1, 2, 4, 6, 50, and 200), but already the hybrid cumulative-additive / multiplicative-additive structure of the system appears to have existed.[1] Most of the signs are recognizably ancestral to those of a somewhat later date, such as the more complete sets of numerals found at Nana Ghat (mentioned above) and at Nasik Cave. While there is no paleographic evidence of Brāhmī numerals prior to 300 BC, some researchers have argued that the Brāhmī numerals were used much earlier. Datta and Singh (1962 [1935]: 37) claim that, because the Aśokan inscriptions are found all over India, the Brāhmī system must have been developed much earlier than the paleographic evidence would indicate, perhaps between 1000 and 600 BC. This spurious use of the 'age-area' method (determining the age of features by their geographical distribution) can no longer be taken seriously. In a situation such as that of the Mauryan Empire, where an enormous region was quickly encapsulated within a single polity, we ought not to be surprised if the administrative inscriptions of that empire are widely distributed. While it was certainly plausible for nineteenth-century Indologists to hope to find earlier paleographic evidence for the numerals, such hopes now seem very remote indeed. On this basis, I am in agreement with Salomon (1996, 1998) and many other Indologists that a mid-3rd century origin for the Brāhmī numerals and script is very probable.

The question of the ultimate origin of the Brāhmī numerals - specifically, whether or not they constitute a case of independent invention, and if not, on which ancestor(s) they were modelled - is unresolved, and is made more complex by the politicization of

---

[1] This opinion contradicts that of Guitel (1975: 605), who sees the Aśoka numerals as being of her type IC' on the basis that the sign for 200 does not sufficiently resemble a ligatured multiplicative 100 x 2. The question remains open, but I am not willing to postulate a significant shift in the principle of a numerical notation system on the basis of a single sign from a handful of inscriptions.

the matter. Because the issue is controversial, I do not pretend that my conclusions can be anything more than tentative, but my approach is quite different from that of earlier scholars. In the past, much weight was given to the paleographic comparison of individual signs. I believe that the consideration of the system's structural features and historical context of origin, supplemented by paleography where appropriate, will be a more fruitful approach. There are at least nine main theories of the origin of the numerals, none of which can be conclusively accepted (although several can clearly be rejected). While it may seem unnecessary to examine and reject numerous theories, some of which have not been tendered seriously in over a century, no other analysis of the origin of Brāhmī numerals brings all these theories together.

1. Brāhmī numerals are derived from letters of the Brāhmī script (Prinsep 1838; Woepcke 1863; Indraji 1876; Renou and Filliozat 1953; Datta and Singh 1962 [1935]; Gokhale 1966; Verma 1971). Prinsep believed the Brāhmī numeral-signs to be acrophonic – taken from the first sounds of the appropriate Sanskrit numeral-words, and Woepcke's important paper promoted this hypothesis. Indraji extended the hypothesis by claiming that the grammarian Pânini (c. 700 BC) used the alphabet with numerical values. A few of the Brāhmī numeral-signs do resemble signs for letters or groups of letters, if one is prepared to accept certain paleographic transformations. As well, many other scripts assign numerical values to phonetic signs, including the alphabetic family (ch. 5) and some later Indian systems (see below). Yet where alphabetic numerical notation systems assign numerical values following the specific letter-order of a script, the Brāhmī numerals do not do so, nor are they acrophonically based on numeral-words. The limited resemblances between the script and numerals are probably due to the convergence of the numerical system and script-signs over time. Bühler (1896), who had earlier supported the theory, abandoned it in favour of the theory of Egyptian origin. Renou and Filliozat (1953) note that in texts containing both purported 'letter-numerals' and the corresponding signs used phonetically, the forms of the two varieties are quite different. While some Indian scholars, such as Gokhale and Verma, continue to defend it, they admit that there no resemblances between many numeral-signs and the script-signs used at the same time, and that much manipulation of the signs is needed to make similarities appear. I reject this theory entirely.

2. Brāhmī numerals are derived from the signs of the Kharoshthi script (Cunningham 1854; Bayley 1882). This theory is similar, except with the signs of the Kharoshthi script in place of Brāhmī used as numeral-signs. That it could even be offered is strong evidence that the paleographic resemblances between the Brāhmī numerals and the phonetic signs of any script are obscure indeed. Cunningham argued for the acrophonic use of phonetic signs for the numerals 4 through 9 (1 through 3 being simple strokes), ignoring the issue of the signs for the tens, hundreds, and thousands. Even these few resemblances have now been rejected, and the theory has not had any supporters for over a century. It must therefore be soundly rejected.

3. Brāhmī numerals are derived from the Kharoshthi numerals (Ifrah 1998). Ifrah offers this theory as one of six he examines and rejects. Both Kharoshthi and Brāhmī numerals were used in South Asia. Furthermore, the Kharoshthi numerals, which are part of the Levantine family and descended from Aramaic numerals, are multiplicative-additive above 100, which may also be true of Brāhmī, as discussed above. In all other respects, however, the systems are entirely different: Kharoshthi numerals are cumulative-additive below 100, use special signs for 4 and 20, and are written in a right-to-left direction. Moreover, there is no good evidence for Kharoshthi numerals before the late 2nd century BC. This theory cannot be seriously sustained, and no scholar has tried to do so.

4. Brāhmī numerals are derived from prehistoric cumulative "tally-mark" signs (Woodruff 1994 [1909]; Salomon 1998; Ifrah 1998). Woodruff held that both the Chinese and Brāhmī numerals derived from a hypothetical ancient set of tally-signs for 1 through 9, in which 1 through 5 were written with cumulative lines and 6 through 9 with complex signs made from the appropriate number of lines. These would then have spread (perhaps from Central Asia) to both China and India prior to the development of the Brāhmī numerals (Woodruff 1994 [1909]: 53-60). Ifrah's theory is similar in nature but very different in scope; he argues that there are "universal constants caused by the fundamental rules of history and paleography" that render likely the theory of independent invention (Ifrah 1998: 390). He argues that a hypothetical cumulative-additive system for writing the numbers 1 through 9 became abbreviated and ligatured into a ciphered-additive system before the time of the first Brāhmī inscriptions. Ifrah's argument is completely unprovable and untestable. It can only be considered as a specific explanation in this one case, rather than as a general and unilinear law of historical progress in numeration. Salomon is far more agnostic; he recognizes the problems involved with many of the other theories I am presently discussing, and simply notes that numerical signs are sometimes "cursive reductions of collocations of counting strokes", citing the hieratic and demotic systems as examples (Salomon 1998: 60). One can hardly disagree with his contention that, with the early history of the Brāhmī numerals perhaps lost forever, their origin might be independent, if no diffusionary argument can be proven.

5. Brāhmī numerals were independently invented with no specific stimulus (Smith and Karpinski 1911; Kaye 1919). Kaye argued that the Brāhmī numerals developed in a specifically Brāhmī context during Aśoka's time, and that the form of the numeral-signs and the structure of the system were entirely separate from any other script or numerical notation system. He believed its structural features represented different stages in the system's development: 1-3 (cumulative unit-strokes) came first, followed by 4-30 (ciphered signs), then 40-300, which he thinks are additive combinations of other signs, then lastly 400 and above, which are multiplicative. Kaye's scholarship is marred by an inexplicable stance against Indian creativity and ability in mathematics. While the multiple-stage model he proposes is not historically valid (early Aśoka inscriptions have numbers as high as 200), the idea that the numerals were independently invented remains attractive to many Indologists today. This theory cannot be completely excluded.

6. Brāhmī numerals are derived from numeral-signs of the Indus script (Sen 1971; Kak 1994). This theory postulates that the Indo-Aryan migration into India resulted in their adoption of the Harappan numerals (ch. 10), and that these eventually developed into the Brāhmī numerals. There are no examples of any writing from India in the enormous period between the latest Harappan inscriptions (around 1700 BC) and the first Brāhmī inscriptions (around 250 BC). Furthermore, there is only limited and conflicting evidence for the nature of the Harappan numerical notation system (Parpola 1994). The fact that the Brāhmī and Kharoshthi scripts also use cumulative unit-strokes for 1, 2, and 3 is no evidence of a historical connection with the Indus civilization, as Sen has suggested. Kak's association of the Brāhmī sign for 10 with the Indus 'fish' sign is not useful, because the 'fish' sign almost certainly did not signify 10 in the Indus script. This theory cannot be sustained on the present evidence.

7. Brāhmī numerals are derived from Chinese numeral-signs (Falk 1993: 175-176). In his recent examination of the Brāhmī script, Falk notes some basic resemblances between the Brāhmī and traditional Chinese (ch. 8) numerical notation systems. When the first Brāhmī numerals appeared in the 3rd century BC, the Chinese used a decimal, multiplicative-additive system whose signs were archaic variants of the later Chinese ones. While Brāhmī is decimal and multiplicative-additive for numbers above 100, it is clearly ciphered-additive for the units and tens. The main paleographic similarity between the two systems is that both use horizontal rather than vertical strokes for the units 1, 2, and 3. The enormous variation in both the Brāhmī and Chinese numeral-signs during this period makes paleographic evidence for a historical connection highly dubious. Moreover, there is no evidence of sustained contact between India and China during this period, and I thus reject this hypothesis.

8. Brāhmī numerals were developed on the model of the Greek alphabetic numerals. This theory has been proposed occasionally, but has never been promoted as a definitive answer to the problem. The obvious reason for envisioning such a connection is the appearance of the Brāhmī script and numerals around the time of Alexander's invasion, and the strong trade ties between Mauryan India and the Greco-Persian kingdoms of Parthia and Bactria starting in the 3rd century BC. Both systems are hybrids, using ciphered-additive and multiplicative-additive notation, and both use a base of 10. This argument would be further strengthened if, like the Greek alphabetic numerals, the Brāhmī numerals were derived from characters of the Brāhmī script (theory 1, above). The Greeks, however, were only beginning to recommence using alphabetic numerals after a hiatus of over a century when the Brāhmī system was developed, and the evidence for the 'alphabeticity' of the Brāhmī numerals is weak at best. There is no paleographic correspondence between the Greek alphabet and Brāhmī numerals. Thus, I reject this theory completely.

9. Brāhmī numerals are derived from the Egyptian hieratic or demotic numerals (Burnell 1968 [1874]; Buhler 1963 [1895], 1896; Salomon 1998). Virtually every Indologist and historian of mathematics who has studied the Brāhmī numerals in the past century has mentioned this theory, first developed by Burnell but promulgated and expanded by Bühler. Burnell argued for direct diffusion from the demotic numerals to Brāhmī, based on paleographic evidence and the systems' contemporaneity. In contrast, Buhler felt that because the hieratic *script* was more like Brāhmī writing than was the demotic, hieratic numerals were the more likely ancestor. Even if Bühler is right about the greater similarity of the hieratic *script* to Brāhmī, there is no reason to believe that the Brāhmī numerical notation system must have a hieratic rather than a demotic origin. Since it is generally agreed that the Brāhmī script is neither hieratic nor demotic but of North Semitic origin, the question of similarity of scripts is entirely separate from that of similarity of numerals. All three numerical notation systems are structurally similar: they are decimal, hybrid ciphered-additive / multiplicative-additive systems. Furthermore, all three represent 200, 300, 2000, and 3000 by adding one or two strokes to the signs for 100 or 1000 in a quasi-multiplicative fashion. Finally, there are some resemblances in around one-third of the sign-forms, and very close resemblances for a few, such as 9 (hieratic = ; demotic = ; early Brāhmī = ) (Bühler 1963 [1895]: 115-119). The similarity in sign-forms is insufficient on its own to demonstrate a connection, but is suggestive in conjunction with other evidence. While the question of cultural contact is trickier than for the hypothesis of Greek origin, Ptolemaic Egyptian traders reached as far as the Malabar Coast, in particular the city of Muziris (modern Cranganore, in Kerala). Additionally, Aśoka is known to have sent Buddhist missionaries to Alexandria during his reign (Basham 1980: 187). Ifrah (1998: 389) glosses over the theory of demotic origin (as with the demotic numerals in general) and considers only the hypothesis of hieratic origin, which he correctly rejects on the basis that the hieratic numerals were mostly extinct by the time of the invention of Brāhmī numerals. Yet the demotic numerals were used until the late Roman period, and were employed for most commerce in the Ptolemaic period.

In short, while Salomon (1998: 60) correctly notes that we do not have definitive evidence to close the case of the origin of the Brāhmī numerals, I believe that a demotic origin should be adopted as a working hypothesis. Of course, the demotic and Brāhmī systems are not identical in either numeral-signs or structure. While the demotic numerals are fully multiplicative only at 10,000, the Brāhmī numerals use multiplication more regularly starting at 100. In addition, demotic numerals are written from right to left, while Brāhmī numerals were written from left to right. Thus, independent invention cannot be excluded completely, with or without the presence of some sort of tally-mark or cumulative-additive system for 1 through 9. Moreover, if diffusion from Egypt is responsible, it remains plausible that the system has a hieratic rather than a demotic origin. More research on this very important subject is clearly desirable.

The Brāhmī numerals, like the Brāhmī script, spread throughout the Indian subcontinent during the Mauryan period and later. Only in the northwest, where Kharoshthi numerals predominated, did the Brāhmī numerals fail to penetrate until around the 4th century AD. They were used primarily for writing dates on stone inscriptions and copper land grants. This is fortunate, because it means that a full set of numeral-signs at least up to 1000 can be derived from the relatively high numbers being expressed, while at the same time the numeral-signs can be assigned an exact date. Other functions for which Brāhmī numerals were used include stichometry and the recording of financial transactions. While it is interesting to speculate on the use of Brāhmī numerals on other materials than stone and copper, the unsuitability of the Indian climate and geography for perishable materials to survive renders such ideas untestable. Yet many manuscripts from Central Asia have survived that contain a variant of the Brāhmī numerals (e.g. texts in the Tocharian language). Similarly, while the presence of a great mathematical tradition can be inferred on the basis of later manuscripts, these use the modern ciphered-positional numerals or the unusual alphasyllabic systems (see below); there is no surviving evidence that the Brāhmī numerals were used for arithmetic or accounting.

After the Kharoshthi script died out in the 4th century AD, Brāhmī numerals were the only ones used in India until the late 6th or early 7th century, and they continued to be employed for several centuries thereafter. They spread not only throughout the Indian subcontinent, but also throughout Central and Southeast Asia, regions that were heavily influenced by India during this period. There was evidently enormous variation in the shapes of the numeral-signs from location to location, which suggests that readability by a large proportion of the population was *not* a primary goal of the writers. In fact, in some Central Asian manuscripts, numeral-phrases were written from top to bottom rather than from left to right (Renou and Filliozat 1953: 702). The primary regional division, between northern and southern systems, began as early as the 2nd century AD, and these two basic variants diverged further in later centuries.

The end of the traditional Brāhmī numerals was a gradual process, instigated not by external influences but by the invention of ciphered-positional notation beginning in the late 6th or early 7th century AD. This entailed the invention of a zero sign and ridding the system of the individual signs for the decades, hundreds, and thousands. Over the next couple of centuries, the older ciphered-additive forms became increasingly rare and by the 9th century AD, the Brāhmī numerals had been replaced by the modern ciphered-positional system throughout India and Southeast Asia. Only in the southern tip of India and Sri Lanka were additive systems retained (though in an altered form) until significantly later.

From addition to position

The development of ciphered-positional numerals has been held by some to be one of the most important achievements of humanity. No numerical notation systems have been as studied and discussed as the ciphered-positional systems that developed in India, primarily because, through the intermediary of the Arabs, these systems are ancestral to our own Western numerals. Ciphered-positional systems have the undeniable advantage that they can express any number, no matter how large, using a small set of numeral-signs. While it would be teleological to portray the history of numerals in a linear fashion leading to our own system, we certainly need to explain why the ciphered-positional systems originating in India eventually came to be used almost universally.

At some point, probably in the early 7th century AD, a change began in the writing of dates in the Brāhmī-derived scripts of India and Southeast Asia. Instead of writing lower numbers with ciphered-additive notation and higher numbers with multiplicative-additive notation, all numbers were written using paleographic variations of the nine Brāhmī numeral-signs and a dot to indicate zero. This system was **ciphered-positional**, because the only indication of a sign's exponential multiplier is its position within a numeral-phrase. Of course, by this time various Brāhmī-derived scripts were already in existence throughout India and Southeast Asia. The spread of the older additive systems (between the 3rd century BC and the 7th century AD, depending on the region) was then followed by a second wave of diffusion of the positional principle and zero (7th century AD onward), in which additive systems change to positional ones. This idea is not as odd as some (e.g. Datta and Singh 1962 [1935]: 39) seem to think, because the change is actually a very simple one. The units are retained in their form, but the other signs (for the decades, hundreds, etc.) are discarded, and replaced with a single sign for zero. This process is confirmed by comparing the signs from 1 to 9 in the non-positional (Brāhmī-derived) systems with the very similar unit-signs used with a zero in later periods.

It is true that this system did not represent the first invention of the zero; the Babylonians and the Maya already used a zero and the positional principle by this time. In fact, it has been argued by many of an older generation of scholars that the Babylonian zero diffused eastward to India just as it diffused westward to Greece (Février 1948: 585; Menninger 1969: 398-9). Yet the Babylonian and Maya systems were both cumulative-positional, and used a sub-base in addition to a base (Babylonian 10 and 60, Maya 5 and 20). Thus, there is no historical relation between these two systems and the later Indian one.[2] The Indian positional system is unique in that it combines ciphering, a zero, and a single, decimal base.

It is frequently claimed that the earliest example of ciphered-positional numerals is found on the Sankheda or Mankani copper plate bearing the date 346 in the Kalacuri

---

[2] Seidenberg's (1986) assertion that the Maya zero came from India across the Pacific Ocean is quite preposterous.

era, which translates to 595 AD (Bühler 1896: 78; Smith and Karpinski 1911: 46; Das 1927: 118; Kak 1990: 199; Ifrah 1998: 401). This plate is a donation charter of Dadda III, used to specify a land grant. When discussing any land grant, the issue of a later forgery always arises, as attempts to claim land by producing such evidence were common in India (as elsewhere). Since a significant amount of our paleographic evidence for early positional numerals comes from such land grants, we must be cautious to avoid dating inscriptions simply by the date as inscribed, but also take paleography and historical context into account. We must also remember that texts containing positional notation that are transcriptions or translations of earlier works must not simply be assigned an early date based on their putative earlier authors.[3] Ifrah believes the paleographic form of the numeral-signs to be unquestionably that of the 6th century AD. He notes that the most prominent of the early scholars claiming forgery, G.R. Kaye (1919: 346), claimed that *all* positional numerals in India prior to the 9th century AD were forgeries, in part due to his prejudice that Western numerals could not possibly be descended from those of the Hindus. While such claims are excessive, Indian epigraphy has progressed mightily in the past seventy-five years, and Indologists still seem very wary of the Sankheda plate (Salomon 1998: 61). The fact that it is ninety years older than any other inscription mentioned by Ifrah suggests that we need to question carefully any 6th century AD evidence for ciphered-positional numerals in India.

The earliest surviving and unquestionable examples of ciphered-positional numerals with a zero derive, not from India itself, but from Southeast Asia, in Khmer, Old Malay, and Cham inscriptions from the late 7th century AD. A calendrical inscription found at Sambor (in Cambodia) and written in a mixture of Old Khmer and Sanskrit is dated 605 in the Çaka dating system, or 683 AD (Coedès 1931: 327). In this inscription,

---

[3] In a most egregious case, the Bakshali mathematical manuscript, which consists of seventy leaves of birch-bark, has been attributed dates from 200 to 1200 AD, but on paleographic evidence (including the presence of many positional numeral-phrases) is probably towards the latter end of this enormous range (cf. Mukherjee 1977; Smith and Karpinski 1911: 43; Struik 1948: 67).

the zero appears as a small dot; this is the first positional zero known from South Asia[4]. As the Sambor inscription is a calendrical passage rather than a land grant or financial document, it is unlikely to be a forgery (Diller 1996: 126). Similar inscriptions from the Old Malay kingdom of Sriwijaya have been found at Palembang and at Kotakapur on the nearby island of Bangka dating to 683, 684, and 686 AD, or 605, 606, and 608 Çaka, respectively; in these cases, the zero was written as a circle rather than as a closed dot (Diller 1995). As Diller notes, it is intriguing that the Old Khmer and Old Malay inscriptions appear in the same year (1995: 66). Nevertheless, there is no reason to believe, as Kaye (1907) did, that the existence of these inscriptions must mean that ciphered-positional numerals actually originated in Southeast Asia and diffused from there to India. The existence of intermediate additive-positional forms from the 6th and early 7th centuries AD, which I will discuss below, coupled with the probability that *some* of the earlier copper grant plates are authentic, make it likely that the invention of ciphered-positional numerals occurred 50 to 100 years earlier. The exact location of its development cannot be pinpointed.

Several authors claim that ciphered-positional numerals existed long before the 7th century AD in India, but that the evidence for their use has been lost. Datta and Singh make the unusual argument that, because the 8th century AD ciphered-positional numerals are found on documents that are of a highly conservative style (grant plates, treaties, etc.), they must have been in use for centuries prior to this time in other contexts (Datta and Singh 1962 [1935]: 49-51). Using analogies from Greece and the Arab world, in which numerical notation systems took five to eight centuries to achieve popularity, they argue that ciphered-positional numerals must have been invented between the 1st century BC and the 3rd century AD. Even if such a spurious historical technique were

---

[4] It has often been wrongly claimed that an inscription found at Gwalior (dated to 876) is the first South Asian inscription with a zero, primarily because this assertion was prominently made by Smith and Karpinski (1911: 52).

valid, the 500-800 year figure in the Greek and Arab cases represents the time from the first surviving appearance of each system to when it became predominant. We cannot simply tack several centuries of unattested use onto the history of every system.

A more compelling argument for an early origin of the positional principle and the concept of zero comes from literary evidence. Perhaps as early as the Vedic period, but certainly by the 4th century AD, special cryptic numeral-words for one through nine could be combined in positional fashion with a word for zero (most commonly *sunya*, or "emptiness, void") in order to represent dates verbally in a fashion quite different from ordinary Sanskrit number-words (Datta and Singh 1962 [1935]: 53-63). These cryptic dates are similar in principle to the *chronograms* used by Hebrew and Arab scholars (ch. 5). Even more notably, the earliest Sanksrit word for the dot or circle for zero, *sunya-bindu* (literally 'void-dot'), is first used in Subhandu's *Vasavadatta*, written around 600 AD (Sen 1971: 175; Salomon 1998: 63). The subject is too complex to cover here, but it does suggest a correspondence (possibly a causal one) between the early use of numeral-words and the structurally identical later use of the ten numeral-signs. Even so, we do not need to accept a date of origin for ciphered-positional numerical notation that is significantly earlier than the epigraphic evidence would indicate.

I accept, along with Salomon (1998) and Ifrah (1998), that a literary origin of the concept of a "zero-space" in Hindu thought, the use of chronograms, and the term *sunya-bindu* in the 5th and 6th centuries AD may have prefigured the eventual development of ciphered-positional numerals. If so, then the invention of the zero may not have a mathematical origin, as is often supposed, but rather a religious or literary one (although religious and mathematical thought are not entirely unrelated in the Hindu tradition). Almost all the attested early ciphered-positional numerals are decidedly non-arithmetical, and are simply used to register dates and other numbers on inscriptions and copper plates. While this may be an artifice of the differential survival of these materials vis-à-vis mathematical manuscripts, it would be erroneous to assume an arithmetical function for the numerals and then to use this assumption to hypothesize an ancient mathematical tradition of ciphered-positional numerals.

Any claim that ciphered-positional numerals were used prior to the middle of the 6th century AD appears to be patently false, and any claim prior to the middle of the 7th century AD requires careful examination. The Indian climate and topography are not particularly suitable for the survival of materials other than stone and metal, and we certainly do not have as much evidence as we would like. Nevertheless, there is plenty of inscriptional evidence for use of the old additive numerals from the 6th through the 8th centuries AD, declining significantly only in the 9th century. To accept that all the evidence for ciphered-positional numerals was lost where so much survives for the additive system is simply preposterous.

A significant number of inscriptions dating from the late 6th to the middle of the 8th century AD from the Orissa region are written with unusual mixed structures combining the features of the older additive and newer positional notations (Datta and Singh 1962 [1935]; Acharya 1993; Salomon 1998). The earliest of these appears to be from the Urlam copper plates of the Eastern Gangâ king Hastivarman, dated to 578 AD, in which the Gangâ era year 80 is written as the additive sign for 80 followed by a zero, but this date may be questionable (Salomon 1998: 62). Acharya (1993) describes many Orissan inscriptions dating from 635 to 690 AD in which dates such as 137 are written as '100 3 7' rather than '137'. This series of dates leads directly into the first fully positional date found in India, on the Siddhantam grant of Devendravarman (195 Gangâ = 693 AD), just ten years after the Southeast Asian examples mentioned above (Salomon 1998: 62).[5] Datta and Singh (1962 [1935]: 52) mention some additional 8th century AD examples combining additive and positional notation. Datta and Singh characterize these hybrids

[5] I am uncertain what to make of Mukherjee's (1993) assertion that the copper-plate inscription of Devakhadga expresses the date 73 in the Harasha era (starting 606 AD) using positional numerals, which would thus be dated to 679 AD.

as representing the gradual forgetting of the older system by writers. Their argument rests on the claim that these are quite late examples of additive notation and that the positional principle was well-established by this time, which is supported only by their odd technique of inferential reconstruction. I find it far more likely that they represent incomplete attempts to incorporate the novel positional principle into inscriptions. These mixed numeral-phrases confirm the hypothesis of a 7th century origin of positional numerical notation in India. Even so, I do not believe that we have enough evidence to conclude from this, as Acharya (1993: 58) does, that the development actually occurred in Orissa, since it might have occurred slightly earlier elsewhere.

The 8th century AD provides considerable evidence for the additive system; however, the positional system gained significant ground, and by the end of the century, the new system was preferred. Around this time, the spread of scientific knowledge from India to China (primarily through the medium of Buddhist scholarship) led to awareness of the ciphered-positional numerals in China. In the Khai-yuan period (713 to 741 AD), the Indian astronomer Qutan Xida (Gautama Siddhartha) translated an Indian calendar into Chinese, using positional numerals (with a dot for zero), and commented on their ease of use (Gupta 1983: 24). In the 9th century AD, the additive Brāhmī system becomes much scarcer. Salomon (1998: 62) notes a striking late example from the Ahar stone inscription in north-central India, which is a composite record of documents of different dates; those up to 865 AD are all dated using additive numerals, and those from 867 AD with positional numerals, providing precise information on the date of replacement. The plate of Vināyakapalā (931 AD) is an extremely late northern (Nagari) inscription containing the older system (Singh 1991: 170). By the 10th century AD, only the far south of India (Tamil and Malayalam-speaking areas) consistently used additive systems, but even there, the old system was replaced by a purely multiplicative-additive structure. Of all the descendants of the Brāhmī system, only the Sinhalese numerical notation system preserved the old ciphered-additive / multiplicative-additive structure until comparatively recently.

Modern South Asian

After the 9th century AD, the transformation of Brāhmī numerals into modern ciphered-positional forms with a zero was complete. The structural evolution of the Indian systems ended at this point, although paleographic developments in the numeral-signs continued to the present day. It is well beyond the scope of this work to describe in detail the enormous amount of paleographic data concerning the development of Indian numeral-signs from 800 AD to the present day (cf. Salomon 1998; Ifrah 1998: 367-385 for more complete analyses of this issue). Nevertheless, a look at some of the more important variations on this common pattern of ciphered-positional decimal systems is warranted, particularly for systems still in use today (or in the recent past). Many of these systems (or very close descendants thereof) have been employed for well over one thousand years and continue to be used. Most major South Asian languages and ethnic groups have their own alphasyllabaries and sets of numerals. While their numerical notation systems are structurally identical to one another (and to Western numerals), they have not been replaced completely by the systems of dominant neighbouring or colonial states, but neither have they diffused beyond a limited region. Today, all these indigenous systems are in competition with Western numerals, especially for commercial and scientific purposes. In religious and formal contexts, the traditional numerals are still strongly preferred.

North India

The ancestor of the northern Indian numerical notation systems is the Brāhmī system used in the Gupta Empire, which ruled most of northern India from the Indus to the Ganges from the 4th to 6th centuries AD, and influenced most of northern and central India. At this early period, Gupta Brāhmī numerals were non-positional, but the idea of positionality and the zero sign spread quickly through the systems of the region. The most common modern varieties of this sub-family (all ciphered-positional and decimal) are the Bengali, Devanagari, Gujarati, Marathi, Oriya, Nepali, and Punjabi; they are thus used in a swath across Pakistan, northern and central India, Nepal, and Bangladesh. The northern Indian systems are also directly ancestral to both the modern Arabic and *ghubar* numerals associated with the Arabic script, and thus, indirectly, to our own Western numerals. The similarities between Western numeral-signs and many of the north Indian numerals, especially for 0, 2, and 3, are quite evident in Table 6.2.

Table 6.2: North Indian numerical notation systems

Script	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Bengali	০	১	২	৩	৪	৫	৬	৭	৮	৯
Devanagari	०	१	२	३	४	५	६	७	८	९
Gujarati	૦	૧	૨	૩	૪	૫	૬	૭	૮	૯
Marathi	०	१	२	३	४	५	६	७	८	९
Oriya	୦	୧	୨	୩	୪	୫	୬	୭	୮	୯
Punjabi	੦	੧	੨	੩	੪	੫	੬	੭	੮	੯
Nepali	०	१	२	३	४	५	६	७	८	९

Central Asia

The Gupta script also gave rise to a small number of scripts in the Himalayas and Central Asia, of which the most important are the Mongolian and Tibetan. The Tocharian script had used a variant of the Brāhmī additive numerals from the 6th to 8th centuries AD, but the Tocharian language and script died out before the introduction of positionality. Tibetan writing and numeration developed in the 9th century, and Mongolian developed from Tibetan in the 13th century, so neither of these systems had non-positional antecedents.[6] It is obvious that these systems are related to the northern Indian systems. The classical Mongolian numerals were usually written from top to bottom in vertical columns, but the forms listed here are those used when they were written from left to right. These systems are shown in Table 6.3.

Table 6.3: Central Asian numerical notation systems

Script	0	1	2	3	4	5	6	7	8	9
Tibetan	༠	༡	༢	༣	༤	༥	༦	༧	༨	༩
Mongolian	᠐	᠑	᠒	᠓	᠔	᠕	᠖	᠗	᠘	᠙

South India

The scripts of the southern half of the Indian peninsula diverged from those of the north as early as the 2nd century AD. There are five modern scripts in this family: Telugu and Kannada, two closely related scripts of east-central India, along with Tamil, Malayalam, and Sinhalese. All of these are derived from the Bhattiprolu script, used around the same time as the Gupta script in the north. Of these five, only the numerical notation systems of Telugu and Kannada are regularly ciphered-positional; Tamil and Malayalam are multiplicative-additive[7], while Sinhalese retains the hybrid cumulative-additive / multiplicative-additive structure of Brāhmī. The Telugu and Kannada signs are much closer to the other ciphered-positional systems of South Asia (Telugu is very close to the North Indian systems, for instance) than they are to the three non-positional systems. This shows that the development of numerical notation systems does not simply

[6] Despite Ifrah's assertion (1998: 382) that each of the Agnean, Kutchean, and Khotanese scripts of Chinese Turkestan would have used a set of ten positional numerals, I know of no evidence that this was the case.

[7] Malayalam is tending towards the increased use of ciphered-positional notation with a zero in the modern era.

follow paleographic changes in scripts. Certain paleographic changes in the numeral-signs probably arose around the time of the adoption of the positional principle in Telugu and Kannada, a transmission that did not occur in the other three systems. The Telugu and Kannada numerals are shown in Table 6.4, while the other three systems are described later in this chapter.

Table 6.4: South Indian numerical notation systems

Script	0	1	2	3	4	5	6	7	8	9
Telugu	౦	౧	౨	౩	౪	౫	౬	౭	౮	౯
Kannada	೦	೧	೨	೩	೪	೫	೬	೭	೮	೯

Southeast Asia

As already mentioned, Southeast Asia, far from being a cultural backwater or simple recipient of positional notation, may be the birthplace of ciphered-positional numerals. Scripts such as Kawi (the ancient script of Java) and Cham (used in Vietnam until the 13th century) originally used hybrid numerical notation systems on the Brāhmī model, but these began to transform into ciphered-positional systems in the 7th century. The modern descendants of these systems include Khmer, Thai, Burmese, Lao, Balinese, and Javanese. Of these, Balinese and Javanese are closely related to one another but paleographically distant from any other South Asian systems. They use Javanese letters to represent certain numbers, while retaining older signs derived from Kawi for the others (0, 4, 5, and 6). The Southeast Asian systems are shown in Table 6.5.

Table 6.5: Southeast Asian numerical notation systems

Script	0	1	2	3	4	5	6	7	8	9
Khmer	០	១	២	៣	៤	៥	៦	៧	៨	៩

Thai	๐	๑	๒	๓	๔	๕	๖	๗	๘	๙
Burmese	၀	၁	၂	၃	၄	၅	၆	၇	၈	၉
Lao	໐	໑	໒	໓	໔	໕	໖	໗	໘	໙
Balinese	᭐	᭑	᭒	᭓	᭔	᭕	᭖	᭗	᭘	᭙
Javanese	꧐	꧑	꧒	꧓	꧔	꧕	꧖	꧗	꧘	꧙

Tamil

The Tamil script and numerical notation system are derived ultimately from Bhattiprolu, the southern variety of the Brāhmī script that developed in the 1st or 2nd century AD. Its immediate ancestor is the Grantha script, which is ancestral to Tamil, Malayalam, and Sinhalese but not to other southern Indian scripts such as Kannada and Telugu. It is used primarily in the far southeast of India as well as parts of Sri Lanka. The Tamil script is alphasyllabic and similar to other Brāhmī-based scripts, but has unique features, such as the ability to represent consonant clusters as a sequence of individual consonant signs rather than using a single sign for several sounds. Similarly, the Tamil numerical notation system is rather different from those of other Brāhmī-derived scripts. The Tamil numeral-signs are shown in Table 6.6 (Guitel 1975: 614-15).

Table 6.6: Tamil numeral-signs

	1	2	3	4	5	6	7	8	9
Units	௧	௨	௩	௪	௫	௬	௭	௮	௯
10	௰		100	௱		1000	௲		

The numeral-signs are derived ultimately from those used in Brāhmī, and are thus related to all the systems of India and Southeast Asia. Following the Indian pattern, numeral-phrases, like the script itself, are written and read from left to right. The Tamil numerals are *not* identical in structure to the older Brāhmī numerals, as has sometimes been claimed (Smith and Karpinski 1911: 52). The traditional Tamil system is multiplicative-additive and decimal; thus, 6408 would be written as ௬௲௪௱௮ (6 x 1000 + 4 x 100 + 8). There is no exponent multiplier for the ones. The structure of Tamil numeral-phrases for 10,000 and higher presents certain typological complexities. Tamil has no signs for 10,000 or higher exponents of 10; nevertheless, large numbers were expressed by placing an appropriate numeral-phrase before the sign for 1,000, then multiplying. Thus, 800,000 would be written as ௮௱௲ (8 x 100 x 1000). There is no ambiguity in this phrase's meaning, because phrases are always read strictly from left to right.[8] This is the only instance where a lower exponent sign may precede a higher one.

The Tamil numerals acquired their distinct structure in the medieval era, perhaps in the 8th century AD, although it is not clear exactly when the divergence arose. Since the Tamil and Malayalam systems are structurally identical to each other, and both are descended from the Grantha script, perhaps the Grantha script also used multiplicative-additive notation. The change from hybrid ciphered-additive / multiplicative-additive to purely multiplicative-additive structure is easily accomplished; because Brāhmī numerals are multiplicative above 100, all that is required is that the nine individual signs for the decades 10-90 be replaced by a single sign for 10. At this early period, we know them largely from inscriptions on stone, although we cannot exclude the possibility that they were used in other contexts. The numeral-signs are derived from those of the Grantha script, and are closely related to others of southern India. It has sometimes been claimed that the Tamil numerals are a uniquely Dravidian invention using letters of the alphabet, and, indeed, there are resemblances between the numeral-signs for 1-9 and nine Tamil phonetic signs (Burnell 1968 [1874]: 68; Ifrah 1998: 372). Nevertheless, I agree with Ifrah and most other scholars that, since these resemblances can only be found by comparing the modern paleographic forms of the numbers and letters, this argument cannot be

---

[8] Curiously, this system is structurally identical to the Armenian alphabetic notation of Anania Shirakatsi (ch. 5), but it would be an error to make too much of this resemblance.

offered as a theory of their origin. Rather, the similarity is probably due to a later assimilation of the numeral-signs to the phonetic signs.

It is interesting to speculate why only Tamil and Malayalam, of all the South Asian systems, altered the Brāhmī ciphered-additive / multiplicative-additive system to a purely multiplicative-additive one (Sinhalese retained the older structure for many centuries, while all other Brāhmī-derived systems became ciphered-positional). I suspect that the persistence of Buddhism in southern India and Sri Lanka under the Cholas, after most of the rest of India had adopted Hinduism - as well as ciphered-positional numerals - may be a partial explanation. In addition, the Chinese traditional numerals are multiplicative-additive, so contact with Chinese Buddhists might have stimulated the development of the unique notations of southern India, or, more likely, made their retention more appealing. Because there is no paleographic similarity between the Chinese and Tamil numerals, I do not believe that diffusion from China was involved. Of course, none of this evidence explains why the Tamils retained their system even after adopting Hinduism.

At some point in the system's history, an abbreviated form of the Tamil numerals developed that, for some numbers, adds an element of positional notation by omitting the exponent-signs for 10, 100 and 1000. For instance, Pihan notes that while 21 was traditionally written ௨௰௧ (2x10 + 1), it could also be written ௨௧, abbreviating the phrase without any loss of information (Pihan 1860: 117). Such numeral-phrases appear to be purely ciphered-positional. Of course, while this presents no problems for numerals that lack any empty positions, a zero sign is needed in other cases; however, no zero appears in any Tamil writings before the 20th century. Sometimes, rather than using a sign for zero, Tamil writers used the exponent-signs for 10, 100 and 1000 to eliminate ambiguity. Guitel cites one instance where 2205 is written as ௨௨௱௫ (2,2,100,5), which indicates that the second 2 is to be understood as a hundreds value rather than as a tens value, and that therefore the first 2 must be understood as 2000 (Guitel 1975: 614-15).

Such phrases combine multiplicative-additive and ciphered-positional notation in an intriguing way that belies the idea that we can assign a single typological label to the later Tamil system.

These mixed multiplicative and positional phrases are no longer used, and appear to be a product of the colonial period, when contact with the West began in earnest. Today, some formal Tamil writings use the traditional numerals, while for most commercial and informal purposes an ordinary 0 sign is used, making the system ciphered-positional. Of course, most literate Tamils are familiar with and use the Western numerals. The survival of such an ancient and peculiar system under conditions of long cultural contact and domination by users of ciphered-positional systems such as the French, British, and other peoples of India, is quite remarkable.

Malayalam

The Malayalam script, like Tamil, is derived from the Grantha script of southern India. It is used to write the Dravidian language of the same name used in Kerala at India's southwestern tip. It first emerged as a distinct script around 700 AD, although its letters and numeral-signs are closely related to those of the other Brāhmī-derived scripts. The Malayalam numerical notation system, like Tamil and Sinhalese, escaped the monotonising effect of the spread of Hinduism and the political influence that rendered the northern Indian numerical notation systems structurally identical and paleographically similar. The traditional Malayalam numeral-signs are indicated in Table 6.7 (Pihan 1860: 122-125; Ifrah 1998: 335).

Table 6.7: Malayalam numeral-signs

	1	2	3	4	5	6	7	8	9
Units	൧	൨	൩	൪	൫	൬	൭	൮	൯
10	൰		100	൱		1000	൲		

The similarities between the Tamil and Malayalam systems are striking. Both systems are decimal and multiplicative-additive, and written from left to right. Thus, one would write 8420 as ൮൲൪൱൨൰ (8 x 1000 + 4 x 100 + 2 x 10). There are many paleographic similarities between the numeral-signs of the two systems, thus refuting the claim that the Tamil numerals are alphabetic in origin. As in Tamil, there is no exponent-sign for the units, and the '1' is understood in any numeral-phrase with a units value. Furthermore, Malayalam numbers above 10,000 are expressed through multiplicative combinations of the sign for 1000 with those for 10 and 100, as necessary. None of these similarities is particularly surprising, given the close cultural and geographic proximity of these two Dravidian peoples. The only structural difference between Tamil and Malayalam numeration is that Malayalam numeral-phrases were never expressed using the mixed additive and positional notation that was occasionally used later in the Tamil system's history.

There are few distinctly Malayalam inscriptions that date before 1000, by which time it had already acquired its multiplicative-additive structure. Because the numeral-signs are derived from those of the Grantha script, it is clear that the Malayalam numerals are native to South Asia, although, as with Tamil, we cannot exclude the possibility of some influence from Buddhist China, given the similarities in the structure of the three systems. The fact that Buddhism was maintained longer in the south than in northern India is likely a partial explanation for the difference in structure. A millennium of trade with and domination by other peoples of South Asia, most of whom used ciphered-positional notation, did not affect the integrity of the Malayalam system. Malayalam inscriptions and manuscripts employed this system regularly until the middle of the 19th century, at which time European contact introduced the zero and the idea of positionality. A new sign for zero was introduced (൦), which, when combined with the nine regular unit-signs, produced a regular ciphered-positional system. Today, the older Malayalam system is used rarely if at all (largely by those who need to understand old texts), and is quickly becoming a historical curiosity.

Sinhalese

The Sinhalese (or Singhalese) script developed from the model of the southern Brāhmī scripts for use among the speakers of Indo-European languages in Sri Lanka, and was clearly influenced by the Grantha script that is ancestral to the Tamil and Malayalam scripts used for writing the Dravidian languages of southern India and northern Sri Lanka. It is an alphasyllabary, written from left to right, and is used today in Sri Lanka and the Maldives. The traditional Sinhalese numeral-signs are indicated in Table 6.8 (Ifrah 1998: 332; Pihan 1860: 140-141).

Table 6.8: Sinhalese numeral-signs

	1	2	3	4	5	6	7	8	9
1s	𑇡	𑇢	𑇣	𑇤	𑇥	𑇦	𑇧	𑇨	𑇩
10s	𑇪	𑇫	𑇬	𑇭	𑇮	𑇯	𑇰	𑇱	𑇲
100	𑇳								
1000	𑇴								

Sinhalese has unique signs for the units, the decades, 100, and 1000. The numeral-signs for the units resemble many of those used in other South Asian numerical notation systems, and many of the signs resemble, but are not derived from, the curved phonetic signs of the Sinhalese script. Numeral-phrases are written from left to right. The system is ciphered-additive below 100, so that 84 would be written as 𑇱𑇤. For the hundreds and thousands, the system is multiplicative-additive, combining the unit-signs with the appropriate exponent-signs; 3684 would then be 𑇣𑇴𑇦𑇳𑇱𑇤 (3x1000+6x100+80+4). It is not clear how the numbers 10,000 and above were written with traditional Sinhalese numerals, though Pihan (1860: 141) speculates that it may have been through multiplicative forms such as those used in Tamil and Malayalam (see above).

The Sinhalese numerals are thus structurally identical to the old Brāhmī system, save for our lack of understanding whether the lower Brāhmī hundreds and thousands were actually multiplicative. The Sinhalese did not adopt the positional system when the peoples of India (except Tamil and Malayalam speakers) and Southeast Asia did, between the 7th and 9th centuries AD. It is interesting to speculate on the effect of Sri Lanka's retention of Theraveda Buddhism, when the rest of India relinquished it. Since the abandonment of Buddhism in India (and the migration of many Buddhists to Sri Lanka) had begun in earnest by the 5th century AD, while the spread of ciphered-positional numerals did not begin until the 7th century AD, this relationship has yet to be demonstrated.

Sinhalese inscriptions and texts used this system throughout the medieval period, seemingly unaffected by the radical changes occurring in Indian numerical notation systems from the additive to the positional principle. Pihan (1860) shows no awareness of any structural changes in the Sinhalese numerals in use at the time he was writing; although his knowledge of the numerals was limited, there is no reason to believe that they were in significant decline in the mid-19th century. Modern Sinhalese writings normally use the Western numerals, although it is my understanding that the traditional numerals are retained for certain formal and religious purposes.

Indian alphasyllabic

The primary numerical notation systems of India were ciphered-additive before the 7th century AD and ciphered-positional afterwards, with only a few systems (Tamil, Sinhalese, Malayalam) remaining additive after that point. The numeral-signs of these systems are abstract and do not resemble closely the letters of the Brāhmī script or its descendants. However, starting around 500 AD, Indian astronomers and astrologers began to use a very different principle for representing numbers: assigning numerical values to the phonetic signs of various Indian alphasyllabic scripts. These systems, known collectively as *varnasankhya* systems, were considered to be distinct from the normal Indian systems that had abstract numeral-signs (Ifrah 1998: 483). Moreover, they were often more flexible than the alphabetic systems of Europe and the Middle East that I discussed in Chapter 5, because they took advantage of the fact that the Indian scripts are alphasyllabaries rather than alphabets or consonantaries. The three systems that I will now discuss - Âryabhata's numerals, *katapayâdi* numerals, and *aksharapallî* numerals, represent an important side branch of the South Asian family. These systems, although used only by a limited group of initiates, are very important for understanding Indian astronomy, astrology, and numerology.

Âryabhata's numerals

The original alphasyllabic numerals, which I will call Âryabhata's numerals after their inventor, are shown in Table 6.9, using modern Nagari signs for convenience (Guitel 1975: 582-583; Fleet 1911a).

Table 6.9: Âryabhata's numerals

क	ख	ग	घ	ङ
ka 1	kha 2	ga 3	gha 4	ṅa 5
च	छ	ज	झ	ञ
ca 6	cha 7	ja 8	jha 9	ña 10
ट	ठ	ड	ढ	ण
ṭa 11	ṭha 12	ḍa 13	ḍha 14	ṇa 15
त	थ	द	ध	न
ta 16	tha 17	da 18	dha 19	na 20
प	फ	ब	भ	म
pa 21	pha 22	ba 23	bha 24	ma 25

कि	खि	गि	घि	ङि
ki 100	khi 200	gi 300	ghi 400	ṅi 500
चि	छि	जि	झि	ञि
ci 600	chi 700	ji 800	jhi 900	ñi 1000
टि	ठि	डि	ढि	णि
ṭi 1100	ṭhi 1200	ḍi 1300	ḍhi 1400	ṇi 1500
ति	थि	दि	धि	नि
ti 1600	thi 1700	di 1800	dhi 1900	ni 2000
पि	फि	बि	भि	मि
pi 2100	phi 2200	bi 2300	bhi 2400	mi 2500

य	र	ल	व	
ya 30	ra 40	la 50	va 60	
श	ष	स	ह	
śa 70	ṣa 80	sa 90	ha 100	

यि	रि	लि	वि	
yi 3000	ri 4000	li 5000	vi 6000	
शि	षि	सि	हि	
śi 7000	ṣi 8000	si 9000	hi 10000	

To explain this system, it is first necessary to add a word about alphasyllabic scripts in India. The basic principle of the Indian alphasyllabaries is that a set of 33 consonant-signs are combined with a set of about 20 diacritic marks that indicate vowels to produce a set of signs for CV syllables; unmarked consonant-signs denote the syllable with the inherent vowel *a*.[9] Thus, while there are many hundreds of possible syllables, to learn the signs of the system one need only learn the two sets of signs, which then can be combined with one another. The 33 unmarked signs, in their assigned order and divided into groups on the basis of similar phonetics (the rows consist of gutturals, palatals, retroflexes, dentals, labials, sonorants, and sibilants, in that order), take on the numerical values 1-25, 30-90, and 100, as shown in the leftmost five columns of Table 14. The ingenious principle involved in this system is that changing the vowel attached to one of the basic signs alters its numerical value. When combined with the vowel *i*, the signs take on the numerical values 100-2500, 3000-9000, and 10,000, as shown in the rightmost five columns. While this means that there are two signs with the value 100 – ha (ह) and ki (कि), this has little potential to cause confusion. Each successive vowel diacritic multiplies the value of the sign by 100 with respect to its predecessor, as shown in Table 6.10 (indicating only the combinations of k + vowels). Using these signs in combination, any number up to 10^{18} could be expressed, and Âryabhata's system by no means exhausts the available diacritics.

[9] The number of consonant-signs and vowel diacritics varies from script to script, and there are also signs for V syllables (isolated vowels) and CCV syllables.

Table 6.10: Order of exponent diacritics

क	कि	कु	कृ	कॢ	के	कै	को	कौ
ka 10^0 (1)	ki 10^2	ku 10^4	kri 10^6	kli 10^8	ke 10^{10}	kai 10^{12}	ko 10^{14}	kau 10^{16}

Numeral-phrases were written with the lowest exponents on the left, which reflects the order of exponents of the Sanskrit lexical numerals, but which is opposed to the Brāhmī numerical notation system, in which the highest exponent was on the left. No sign for zero was needed, and none used. The signs for 11-19 and 21-25 were not strictly necessary; 15 could be written as ङञ instead of ण without any ambiguity, but obviously, the latter was more concise, and similarly 1515 would be written as णणि (15+1500). These extraneous signs normally were not used in numbers such as 85, which was written as ङष (5 + 80) rather than णश (15 + 70). In some cases, these rules were violated (we do not know why), so that in one astronomical table, 106 is written as 16+90 and 37 as 16+21 (Guitel 1975: 587). Table 6.11 indicates several numeral-phrases written alphasyllabically.

Table 6.11: Alphasyllabic numeral-phrases

Value	**Alphasyllabic representation**	**Transcription and sign-values**					
62	खव	kha 2	va 60				
116	तह or तकि	ta 16	ha 100	OR	ta 16	ki 100	
9800	जिसि	ji 800	si 9000				
70,040	रछु	ra 40	chu 70,000				
232,221	पफिबु	pa 21	phi 2200	bu 230,000			
765,432	खयघिलिचुशु	kha 2	ya 30	ghi 400	li 5000	cu 60000	śu 700,000
98,206,025	मविनुजृसृ	ma 25	vi 6000	nu 200,000	jri 8,000,000	sri 90,000,000	
40,000,220,000	फुघे	phu 220,000	ghe 40,000,000,000				

The best way to conceive of this system is as a base-100, or centesimal, multiplicative-additive system with a decimal and ciphered sub-base. Unlike most multiplicative-additive systems, however, there can be up to two unit-signs within each exponent of 100, each of which combines with its own exponent-sign. For instance, in the representation of 9800 in Table 6.11, the signs for 8 (ज) and 90 (स) each combine separately with the diacritic sign for 100 (ि). The system is slightly irregular below 100 in that the basic 33 signs include signs for 11 through 25. An alternate way to classify this system would be as a base-10 ciphered-additive system that is multiplicative above 100, but this is not as satisfactory, because it does not adequately indicate that the exponents of 100 (100, 10,000, 1,000,000...) have distinct exponent signs while other exponents of 10 (1000, 100,000...) do not. This principle was clearly understood by Âryabhata, who distinguished the set of centesimal exponents, or *varga*, from the intermediate exponents, or *avarga* (Das 1927a: 110). It is entirely incorrect to argue, as some researchers have done, that this system is positional, since placing an unmodified consonant-sign in the middle of a numeral-phrase would render it meaningless.

While it is sometimes claimed that the Indian grammarian Pânini used alphasyllabic numerals in the 7th century BC (Datta and Singh 1962 [1935]), this is a highly dubious proposition given the lack of attested writing in India between the end of the Harappan civilization and the rise of the Mauryan Empire. There is no evidence for the system described above, or any other alphasyllabic numeration in India, until about 510 AD, near the end of the dominance of the Guptas over India. It was very probably invented by the mathematician and astronomer Âryabhata, in whose works (later named the *Âryabhatiya* by his disciples) it first appears. Âryabhata, who lived in the small town of Kusumapura in modern Bihar, not only became renowned among Indian scholars of the Gupta empire and later centuries, but also was known to Muslim scholars as *Arjabhad* and in medieval Europe as *Ardubarius* (Ifrah 1998: 447). His work focused on astronomy (he is regarded as the first great Indian astronomer) but also contains much pure mathematics, in addition to the cosmological hypotheses from which the exact sciences in ancient India cannot be divorced. His numerals are concise and readable with little training, and while not infinitely extendable, were certainly capable of expressing the very high numbers in his works. If we accept, as I do, that Âryabhata did not know any positional numerical notation system, then he may have developed alphasyllabic numerals because of the insufficiency of the Brāhmī ciphered-additive system for writing large numbers, a task which could be done very concisely with his own system. However, there is no evidence that the calculations that Âryabhata undertook were done directly with these numerals.

It is possible that in addition to the Brāhmī numerals, Âryabhata was familiar with the Greek alphabetic numerals. Âryabhata's work was inspired in part by Greek astronomical writings, and Fleet (1911a), among others, has argued that both Âryabhata's astronomy and his numerals are derived from Greek sources. However, even if he borrowed the general idea of using script-signs as numerals from the Greeks – and there is no definite evidence either way - this does not tell us very much, because the two systems are radically different. Das (1927a: 111-114) has quite effectively shown that the similarities between the Greek system and Âryabhata's are quite superficial. Even if he did know the Greek numerals, they played little role in the invention of his own system, whose idiosyncratic features, such as a base of 100, are not found in other systems, and are probably a consequence of the alphasyllabary itself, with its permutations of consonantal signs and diacritics.

Another possibility, suggested by Ifrah (1998: 450), is that even though his system was not positional, Âryabhata must have had a complete knowledge of ciphered-positional numeration in order to invent his alphasyllabic system. As I discussed earlier, the concept of *sunya* or "emptiness" existed in the 5th century AD and may have prefigured the use of positional numerals in India, but no good evidence survives for an actual ciphered-positional numeration system prior to the 7th century AD, long after

Âryabhata's death. Ifrah's (1998: 450-451) statement that the use and adoption of Âryabhata's system "caused the Indian discoveries of the place-value system and zero, which took place before Âryabhata's time, to be irretrievably lost to history" is typical of the confused anti-empiricism of recent research on this system. A better theory is that the non-positionality and relative complexity of Âryabhata's system argue against his having been familiar with positionality.

While Âryabhata's numerals were known to Indian astronomers and mathematicians long after his death, they were used solely in the context of commentaries on his work. Otherwise, his system of multiplicative-additive notation fell into disuse. It was replaced, in part, by the regular numerical notation systems of India, but it also gave rise to a variety of successor systems for correlating phonetic signs with numerical values, most notably the *katapayâdi* system. While these successors were not as unusual as Âryabhata's system, they were far more successful, and some continue to be used today.

Katapayâdi numerals

When later scholars experimented with alphasyllabic numeration starting in the 9th century AD, they immediately saw that an alphasyllabary could also be turned into a ciphered-positional system. Known as *katapayâdi,* the signs of this system are shown in Table 6.12 (Fleet 1911b; Datta and Singh 1962 [1935]: 70).

Table 6.12: Katapayâdi numerals

क	ख	ग	घ	ङ
ka 1	kha 2	ga 3	gha 4	ṅa 5
च	छ	ज	झ	ञ
ca 6	cha 7	ja 8	jha 9	ña 0
ट	ठ	ड	ढ	ण

ṭa 1	ṭha 2	ḍa 3	ḍha 4	ṇa 5
त	थ	द	ध	न
ta 6	tha 7	da 8	dha 9	na 0
प	फ	ब	भ	म
pa 1	pha 2	ba 3	bha 4	ma 5
य	र	ल	व	
ya 1	ra 2	la 3	va 4	
श	ष	स	ह	
śa 5	ṣa 6	sa 7	ha 8	

In this system, each V and CV syllable is given a value between 0 and 9. Unlike in Âryabhata's system, changing the vowel of the syllable does not change its numerical value, so that *ka* = *ki* = *ku* = 1. Two of the signs (ña and na) take on the value of zero, as did isolated vowel-signs (those representing a V syllable alone, without any consonantal component), which did not have a numerical value in Âryabhata's system. CCV syllables do not have their own numerical values, but are considered to have the value of the consonant to the left of the vowel, so that *tva* = *va* = 4 and *ntya* = *ya* = 1. As a result, any sequence of syllables can be assigned a numerical value, read with the lowest exponent on the left as in Âryabhata's numerals. Thus, the word *bhavati* or भवति had the numerical value 644. The name *katapayâdi* itself is taken from the four syllables (ka, ta, pa, ya) that are assigned the value 1 in this system. Although it is unusual in that each digit from 0 to 9 has several alphasyllabic values that represent it, structurally this system is an ordinary ciphered-positional and decimal system.

The earliest example of the *katapayâdi* numerals seems to be from the *Grahachâranibandhana* by the astronomer Haridatta in the middle of the 9th century AD (Ifrah 1998: 474). Datta and Singh (1962 [1935]: 71) place its invention around 500 AD and claim that it was known to Âryabhata himself, and their theory is widely accepted today because of van der Waerden's (1963: 55) endorsement. However, there is no clear textual evidence from this period to support this assertion, as far as I can tell. Haridatta was a direct intellectual descendant of Âryabhata, and obviously used his predecessor's system as the basis for his own. Nevertheless, given that the *katapayâdi* is ciphered-positional, like the general Indian positional numerals in ascendance at the time, the existing positional numerals must certainly have influenced him. Thus, this system is very likely a blend of Âryabhata's numerals and the ordinary Indian positional numerals. Aside from being able to express any number, it had the advantage of being able to give any word a numerical value, and given a numerical value, to find many words corresponding to that number. This would have allowed for the construction of various mnemonic devices to aid scholars and students, and would have served a prosodic function (for astronomical texts were written in Sanskrit verse, which had strict metrical rules). The use of *katapayâdi* numerals was also fundamental to the Hindu tradition of number-magic and divination, including chronograms, in which the sum of the numerical values of the signs of a word or verse produced a meaningful date.

The *katapayâdi* numerals, as well as related systems that are identical except for the use of local script-signs and the assigning of different digit-values to various signs, were in continuous use throughout much of India for many centuries. Several variants of the *katapayâdi* developed, most of which change a few numerical values or eliminate the values of certain categories of signs, such as the isolated vowels (Datta and Singh 1962[1935]: 71-72). Some of these systems were unique to one writer, while others were used in specific regions over a longer period. *Katapayâdi* numerals were restricted in use to divinatory, astrological, and astronomical contexts, which is due in no small part to their utility for chronograms. Curiously, they survived much more extensively in southern India (where the additive Tamil and Malayalam numerals were never replaced by positional systems) than in the north (Renou and Filliozat 1953: 708). They were still used in astrological manuscripts and horoscopes in South India even in the late 19th century (Burnell 1968 [1874]: 79-80). Renou and Filliozat (1953: 708) claim that their use in paginating loose manuscripts served a cryptographic function in that the pages of the text, once jumbled, could only be placed back in order by initiates of the system. I know of no definite evidence that the *katapayâdi* system is still used today, though it may be.

Aksharapallî

A third variety of alphasyllabic numerals, sometimes confused with the ciphered-positional *katapayâdi* in the scholarly literature, is known as *aksharapallî* numeration (after *akshara,* the word for the CV syllable-clusters that comprise the basic unit of the Indian alphasyllabaries). Whereas Âryabhata's system was multiplicative-additive, and the *katapayâdi* system was ciphered-positional, the aksharapallî systems are ciphered-additive and decimal, assigning the numerical values 1-9, 10-90, and sometimes also the low hundreds to a set of phonetic signs, but never as high as 1000, to my knowledge. It was used very widely for paginating books, and was written in the margins from top to bottom with the highest exponent at the top.

Unlike the first two alphasyllabic numerical notation systems I have discussed, there was never a single regular system for correlating signs with numerical values in the aksharapallî. Datta and Singh's (1962 [1935]: 73) search through old manuscripts revealed no fewer than three signs for 1, twelve different signs for 4 and nine signs for 60. This is less complex than it seems, because no doubt within each regional tradition, there was a set sequence of signs that would be understood by anyone working within that tradition. In some instances, parts of these sequences may be comprehensible; for instance, in Nepali manuscripts from the 11th to the 14th centuries, the numbers 1 through 3 were represented by the syllables *e, dvi,* and *tri,* which correspond to the Nepali lexical numerals (Burnell 1968 [1874]: 66). In many other cases, though, the signs used appear to have been assigned almost randomly. Datta and Singh (1962 [1935]: 73) list many signs for which they cannot even attach a plausible syllabic value. In some of these cases, dialect differences in pronunciation or paleographic variation among scripts may account for the irregularity of the aksharapallî systems.

The origin of the aksharapallî is quite obscure. They are clearly ciphered-additive, but the lack of a regular correlation between the signs and numerical values suggests no obvious origin. They may very well have originated directly from the Brāhmī ciphered-additive numerals, with only the use of phonetic rather than abstract symbols to distinguish them. This matter is made even more complex by the fact that many modern scholars still maintain that the origin of the Brāhmī numeral-signs was as a modification of phonetic signs (see theory #1, above). Given that Indian thinkers considered the aksharapallî to be part of the *varnasankhya* tradition of alphasyllabic numeration, however, I suspect that they are related to the other alphasyllabic numerical notation systems, although possibly with some influence from knowledge of the Brāhmī system. The primary (indeed, perhaps the only) use of aksharapallî numerals was for the pagination of manuscripts. This may explain in part why there was apparently no need in any such system for expressing numbers in the high hundreds and thousands. Because they were never used for mathematics or arithmetic, there certainly would not have been any reason to abandon them solely on the grounds of efficiency.

Aksharapallî numerals had the greatest and most consistent level of use of any of the alphasyllabic numerals of India. They were used with great frequency in the manuscripts of the Jains until the 16th century, although it is not clear why this system would appeal specifically to Jains (Datta and Singh 1962 [1935]: 74). They also survived for a very long time in Nepal (Burnell 1968 [1874]: 65). Temple (1891) goes into great detail concerning what is clearly a ciphered-additive numerical notation system used for arithmetic by Hindu astrologers in Burma in the late nineteenth century. While he does not indicate whether the signs were alphasyllabic, nor does he use the name aksharapallî (or any other name) to describe this notation, I know of no other ciphered-additive notation likely to have been used in Burma at that time. Aksharapallî numerals appear to have thrived on the Malabar Coast; they were still common enough in Malayalam-speaking regions in the middle of the nineteenth century to be included in some grammars (Bendall 1896). What is common to these groups is that they are relatively distant from the central political and religious movements of India. Their survival in these regions may simply be an artifact of the marginal status of such places in Indian history. Ifrah (1998: 484) indicates that aksharapallî systems continue to be used today throughout India, Bangladesh, Nepal, Tibet, Burma, Cambodia, Thailand, and Java, but it is unclear on what basis he postulates such an extensive present distribution.

Arabic positional

The Arabic script is written from right to left, and is basically consonantal, though with some representation of vowel sounds. The earliest Arab numerical notation system was a hybrid cumulative-additive / multiplicative-additive system immediately descended from the Nabataean numerals (ch. 3); this was replaced after the Islamic conquest of Greek-speaking regions by an alphabetic system (known as the abjad numerals) akin to the Greek, Hebrew, and Syriac systems (ch. 5). Yet as early as the middle of the 7th century AD and certainly by the start of the 9th century, the Arabs were using a ciphered-positional and decimal numerical notation system, of which the modern numeral-signs are shown in Table 6.13.

Table 6.13: Arabic positional numerals

0	1	2	3	4	5	6	7	8	9
٠	١	٢	٣	٤	٥	٦	٧	٨	٩

This system is written with the higher exponents on the left, *not*, as with the earlier Arab numerals, from right to left following the direction of the Arabic script. Thus, 26049 would be written ٢٦٠٤٩. In addition to these signs, there are alternate signs used for 4 (٤) and 5 (٥). The obvious reason why the zero sign would be written with a dot instead of with a circle, like the Western and Indian systems, is that the circle was already assigned the value of five.

While there are certain resemblances between these signs and those used in medieval India, the question of the origin of these numerals will be best answered by examining their early paleographic forms. Table 6.14 compares the Arabic positional numerals found in 11th century mathematical and astronomical treatises with the inscription found at Gwalior, India dated to 876 AD, containing the Nagari numerals used in medieval India. These signs are very similar, and it is thus safe to assume that the Arabic numerals have an Indian origin.[10] In some cases, as for 2, 3, 7, 8, and 9, the Nagari numeral-sign became rotated or inverted, a fact that Ifrah attributes to the practice of some scribes of writing from top to bottom then rotating the manuscript to read it (Ifrah 1998: 532-533). The fact that medieval and modern Arabic scholars are unanimous in attributing an Indian origin to these signs, and that they are called *al-hisab al-hindi* (Indian numerals) merely confirms what is evident from the paleographic evidence.

Table 6.14: Early Arabic and Nagari positional numerals

	0	1	2	3	4	5	6	7	8	9
Arabic	٠	١	٢	٣	٤	٥	٦	٧	٨	٩
Nagari	०	१	२	३	४	५	६	७	८	९

It is not exactly clear when the Indian positional numerals reached the Arab world. In 662 AD, a Syrian Christian bishop, Severus Sebokht, noted the Hindu proficiency in astronomy, commenting that "as for their skilful methods of calculation and their computing which belies description, they use only nine figures" (Nau 1910). The meaning of this statement is unclear, as he does not mention the zero, but it is likely

[10] I am thus entirely unconvinced by arguments such as that of Attié Attié (1975) that the Arabic positional numerals were developed in the Middle East in the 6th century AD and then spread from there to India.

that Sebokht was referring to ciphered-positional numerals; if so, the Muslim world probably would have had some such knowledge as well. Nevertheless, there is no evidence that Christians or Muslims used positional numerals in the 7th or early 8th centuries AD. The first strong evidence for the numerals' transmission to the Arab world is from 773 AD, at which time an Indian astronomer visited the court of the Abbasid caliph Al-Mansur in Baghdad, bringing with him a copy of the *Siddhanta*, a Hindu work of astronomy (Menninger 1969: 410). Al-Mansur had this text translated into Arabic.

Within 50 years, the mathematician al-Khwarizmi wrote his *Arithmetic* (c. 825 AD) using ciphered-positional numerals extensively, prompting later mathematicians and astronomers to follow his lead in replacing the old ciphered-additive "abjad numerals" with the new positional system. While al-Khwarizmi's work does not survive in its original Arabic, we know from later mathematicians, most notably Adelard of Bath's 12th century Latin translation, that al-Khwarizmi not only knew of the numerals but also used them correctly in his own work and advocated their simplicity and functionality. We do not know, however, specifically what numeral-signs he used. The earliest material evidence for positional Arabic numerals comes from an Egyptian papyrus dated 260 A.H., or 873 AD (Menninger 1969: 414). Nevertheless, it seems highly probable that some Arabs, particularly those around Baghdad, knew of them at least by al-Khwarizmi's time, and possibly as early as 775 AD.

From their origins in the late 8th and early 9th centuries AD, the numerals spread throughout the Islamic world reasonably quickly, though not without resistance or confusion. Ifrah (1998: 539-541) provides a number of 10th century examples where conservative scribes and bookkeepers resisted the new numerals in favour of older calculation on the fingers and with numeral-words. Lemay (1977: 440-444) questions the extent to which the Indian numerals were known to the Arabs before the 10th century, and shows that there was confusion among some Arabic scientists over how they worked. Regardless, by the 11th century, positional numerals dominated both in nathematical and non-mathematical contexts throughout the Arab world. They failed to penetrate only its far western regions (North Africa and Spain). There, as early as the 9th century, a second Arabic system of numerals, known as the *ghubar* or "dust-numerals", was used. I will discuss this system below.

The question of the context in which the Arabic numerals arose is a curious one, for we find, contrary to the diffusion of most numerical notation systems, that science, rather than commerce or religion, was the impetus for the transmission of the positional numerals from India westward. The Indians and Arabs shared no common language, religion, or script, and were politically fiercely independent and even rivals. Of course, there was considerable trade and cultural contact between the two regions, and it is possible that the evidence for a commercial origin of the numerals has been lost. From the surviving evidence, though, the initial motivation for the adoption of the positional principle and zero-sign was for the practice of mathematics and astronomy, from which it then spread for other functions.

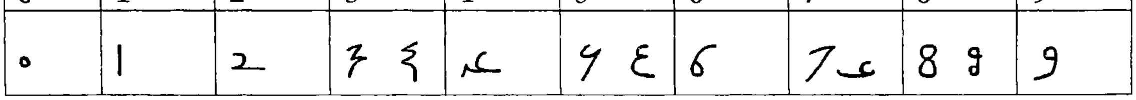
The Arabic numerals enjoy a degree of currency and use in the modern world second only to the Western numerals. They are used regularly in a wide variety of contexts throughout all regions that employ the Arabic script, and are thus found regularly from Morocco to Indonesia. Even so, the rise of global commerce and the effects of mass media have aided the introduction of Western numerals into the Arabic-speaking world, and most literate users of the Arabic script are familiar with the Western numerals. It remains to be seen whether this will have any long-term effects on the use of Arabic numerals, but it seems unlikely at present.

Ghubar numerals

A set of ciphered-positional numerals quite distinct from the regular Arabic system was used in North Africa and southern Spain during the medieval era and sporadically thereafter. These numerals, known as *ghubar* "dust, sand" numerals, are important not only because of their survival in unusual circumstances for several centuries in a peripheral region of the Islamic world, but also because they are the immediate ancestor of the Western numerals. While the paleographic forms of the *ghubar* numeral-signs vary, representative examples are indicated in Table 6.15 (Labarta and Barceló 1988; Gandz 1931; Souissi 1971; Ifrah 1998: 534).

Table 6.15: Ghubar numerals

0	1	2	3	4	5	6	7	8	9



In chapter 5, I showed that North Africa and Spain were quite distinct from the rest of the Arab world, both in their use of a different ordering of the alphabetic or abjad numerals, and in their use of special "Fez numerals". Comparing the ghubar numerals to the standard Arabic positional numerals (either the medieval or modern forms), we can see that while there are resemblances, the two systems differ paleographically. Structurally, they are both decimal, ciphered-positional numerical notation systems written with the highest exponents on the left.

The earliest known examples of the ghubar numerals come from two documents dated from 874 and 888 AD, respectively, in texts from the Maghreb (Gandz 1931: 394). It is perhaps notable that the first textual example of the ghubar numerals comes only one year after the first regular Arabic positional numerals known in Egypt, but this may simply reflect accidents of survival and discovery. Smith and Karpinski (1911: 98), Das (1927b: 359) and Datta and Singh (1962 [1935]) argue that the ghubar numerals are closer to the original Indian forms, and thus are an earlier transmission, than the later Eastern Arab forms. They claim from this that the ghubar numerals were the ones used by al-Khwarizmi and other early mathematicians; however, such a conclusion is overly speculative. Another theory, popular from about 1920 until 1935, holds that the ghubar numerals came from India to Spain via Neo-Pythagoreans in Byzantium, while the standard Arabic numerals came from India via the caliphate of Baghdad (Cajori 1919; Gandz 1931: 395; Miller 1933; Lattin 1933: 184-5). This peculiar theory has no redeeming features, as we have no evidence for ciphered-positional numerals in Byzantine Greece prior to the 12th century (Wilson 1981). Furthermore, it unfairly overstates the Western heritage of the numerals, while limiting their Hindu or Arabic ancestry.

In fact, the ordinary Arabic numerals and the ghubar numerals were quite similar until the 12th century; their numeral-signs for almost all values are similar enough to be explained as graphic variations of a common system of Indian derivation (the medieval Nagari ciphered-positional system). In a 10th century manuscript written by the Persian astronomer Sigzi, the form of numerals used is intermediate between the Arabic and ghubar forms (Mazaheri 1974). Medieval Arabs definitely regarded the two systems differently, even calling them by two separate names, *al-hisab al-hindi* and *al-hisab al-ghubar*, but this does not imply two separate waves of diffusion. Lemay clarifies the situation greatly: "In fact, 'Hindu' and *ghubar* numerals in use among the Arabs belong to one and the same tradition, namely the system of nine symbols and the zero used in value position. Its generic name among the Arabs was *al-hisab al-hindi,* to which the *ghubar* numerals belong" (Lemay 1977: 437).[11] Ghubar numerals are a sub-set of the larger class of Indian-derived numerals, which stand in contrast to the *abjad* numerals described in Chapter 5.

Why, then, did the ghubar numerals survive as a distinct variant in the Maghreb? The term ghubar, with its unusual meaning of "dust" or "sand", has prompted some comment as to the use of the numbers. Das (1927b: 358) and Gandz (1931) agree that this name probably derives from the Hindu and later the Arab practice of using boards covered with dust or sand as calculating boards by drawing figures on them. I believe that much of the variation between the Arabic positional and ghubar numerals can be explained by the differing media and contexts of their original use: the regular system on

[11] Cf. also Ifrah 1998: 529-539 for a good comparison and analysis of the two systems.

tone inscriptions and in texts, and the ghubar numerals for arithmetical calculations on and-boards. Their forms, thus fixed by the separation of contexts, became entrenched hrough centuries of use in disparate parts of the Islamic world.

While the ghubar numerals began as a board-based variant of the Indian numerals, they quickly took on a distinct cultural meaning among the scribes, astronomers, and mathematicians of the western Islamic world. This is no doubt partly because of the relative independence of polities such as the caliphate of Cordoba from the Baghdad-based Abbasid caliphate. Ifrah suggests that the traditionalism of the Maghrebi and Andalusians may partly explain why the ghubar numerals persisted even after the rest of the Islamic world had adopted the modern signs (1998: 539). Regardless of the reason, they were still regularly used in Spain and North Africa in the 15th and 16th centuries, and sporadically thereafter (Labarta and Barceló 1988).

The ghubar numerals would be little more than a paleographic curiosity, merely one of many ways of writing Arabic numerals, if not for the fact that through them, Europe came to adopt ciphered-positional numerals (see below). For the past seventy-five years, it has been essentially agreed by all scholars that the resemblances between ghubar and Western numerals, coupled with the circumstances of the origin of the latter system in medieval Spain, show that the Western numerals derive not from the standard (Eastern Arab) numerals but from the ghubar numerals. The ghubar numerals survived for almost a full millennium, an incredible length of time, considering that the areas where they were used – North Africa and Muslim Spain – were influenced heavily both by Arabic and European cultures. Even more strikingly, they do not appear to have changed greatly in form over that period. Ifrah (1998: 535) provides examples of arithmetical texts written with ghubar numerals from as late as the 18th century, and suggests that the system may have survived into the 19th century, before being completely replaced by the standard Arabic numerals.

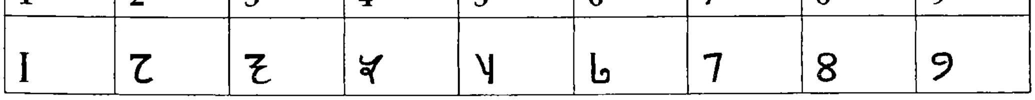
Western

Let us turn finally to the numerical notation system with which we are all familiar, and which is predominant throughout most of the world today: the so-called Hindu-Arabic, or as I call them here, the Western, numerals. From their origin as a foreign and suspicious novelty during the medieval period, these ten signs, unified by the use of the positional principle, have become so familiar that it is easy for the non-specialist to forget that there *are* other numerical notation systems. The ubiquity and universality of the Western numerals make understanding their origin and diffusion all the more important. Unfortunately, no scholarly text has adequately dealt with the topic since Hill (1915), whose work is rather outdated as a result of modern advances in paleography.

The first example of Western numerals is generally held to be the Codex Vigilanus, written in 976 in the monastery of Albelda near the town of Logroño in northern Spain, in which the numerals are described (in Latin) as "Indian figures" (Hill 1915: 29). The nine units are listed, in descending order, but no zero-sign is evident, probably because the signs were intended for use with a counting-board. These signs are shown in Table 6.16.

Table 6.16: Western numerals (Codex Vigilanus, 976)

1	2	3	4	5	6	7	8	9



These figures are very similar to the ghubar numerals shown in Table 6.11, above, and in fact there is no reason to consider them as a separate system, except that they are used in a Latin and Christian text from northern Spain rather than an Arabic one from Andalusia. Lemay has established that Toledo was a major center for the transmission of Arabic knowledge to the Christian West in the 10th and 11th centuries, and believes that later scholars became aware of ciphered-positional numerals through reading Toledan texts (Lemay 1977: 444-5).

These numerals found their way into slightly more widespread usage through the writings of Gerbert of Aurillac (c. 945-1003), who later became Pope Sylvester II in 1000. Gerbert, who had travelled extensively in Islamic Spain, wrote in favour of these numerals in the last decades of the 10th century, advocating their use not in manuscripts but on counting-boards, whose traditional technique of use was similar to that of the bead abacus. He noted that by replacing a large number of tokens placed in any column with a single token bearing one of these signs, calculation became much simpler. No zero-sign was needed because counting-boards are positional by their nature, without the need for a placeholder. These marked tokens, called *apices*, were used by medieval mathematicians, known as *abacists*, between the 10th and 12th centuries, but they never achieved popularity outside this limited group (Evans 1977; Lemay 1977; Gibson and Newton 1995). Hill's extensive study lists fewer than twenty examples of ciphered-positional numerals in Christian Europe prior to the 13th century, of which the majority are from treatises on the *apices*, rather than manuscripts where the numerals are treated as number-signs alone (Hill 1915: 29-31).

The spread of Western numerals into the tradition of manuscript-writing (both in mathematical and other texts) did not really begin until 1202, at which time the mathematician Leonardo of Pisa, better known as Fibonacci, promoted their use in his *Liber Abaci* (Book of the Abacus). Despite its name, the purpose of Fibonacci's text was not to promote the use of the abacus, but rather the use of written numerals for computation, with nine unit-signs and a zero-sign (θ). Later scholars who followed in Fibonacci's wake, such as John de Sacrobosco, who wrote in Paris around 1240, used the term *algorismus* (a corruption of the name al-Khwarizmi) to refer to this new art. Thus, even though he did not use the term himself, Fibonacci was the forerunner of the *algorithmists*, who, in direct conflict with the abacists, promoted the use of written numerals for computation rather than the use of counting-boards (cf. Evans 1977; Murray

978). This technique is the precursor to modern computational techniques with pen and paper.

Despite the unquestionable importance of Gerbert, Fibonacci, and other mathematicians in introducing the ghubar numerals to the West and promoting their use, their eventual adoption is not a vindication of a 'great man' theory of history. The diffusion of the Western numerals from Andalusia and North Africa to the West likely occurred numerous times and by several different routes, some of which were more fruitful than others (Gibson and Newton 1995: 316). Contact between the Arab and Western cultural spheres followed several paths in the early Middle Ages: through Spain, to be sure, but also through Norman Sicily, along main trade routes from African cities such as Tunis and Tripoli to Venice and Genoa, and through the Crusader states.

Far from being an instantaneous adoption, the Western numerals were used only by a small number of Western European scholars until the 16th century. The ordinary populace of Western Europe used Roman numerals, if any, while Eastern Orthodox regions used alphabetic systems such as the Greek or Cyrillic alphabetic numerals. In fact, the use of Western numerals was prohibited in several instances. In 1299, the *Arte del Cambio* or moneychangers' guild of Florence prohibited the use of Western numerals in its registers, a prohibition that was maintained for at least 20 years (Struik 1968). The reason given in the document is to prevent fraud due to the numerals' ease of falsification and confusion resulting from their novelty. Struik rightly notes that socioeconomic explanations for this prohibition may be just as persuasive, taking into account conflict between different factions, some more conservative than others. Since the numerals were a foreign invention, xenophobia and ethnocentrism also may have played a role. Similar prohibitions were enacted as late as 1494 in Frankfurt, where the Bürgermeisterbuch, or mayor's book, instructed bookkeepers not to use Western numerals in performing calculations (Menninger 1969: 427). The true reasons for these prohibitions remain unclear, but the subject is a very interesting one, as it bears directly on the question of how numerical notation systems are transmitted and adopted.

Table 6.17 demonstrates the slow transmission of Western numerals throughout Europe, including both their first occurrence in each region and the period in which they became more commonly known.

Table 6.17: Early Western numerals in Europe

Location	First example	Common use
Spain	976: Codex Vigilanus (Hill 1915: 29)	1490[12]: dating pages in texts
Italy	c. 1050 - 1075: Pandulf of Capua's *De Calculatione* (Gibson and Newton 1995)	c. 1325: banking records and account books in major cities (Struik 1968, Menninger 1969: 428)
France	mid- to late 11th century: abacus treatises (Hill 1915: 29)	c. 1400: dating, accounting, etc.
England	c. 1130: Adelard of Bath's translation of al-Khwarizmi	1525 - 1550: archival records, accounting books (Jenkinson 1926)
Germany	1143: translation of al-Khwarizmi into Latin at Vienna (Menninger 1969: 411)	1525 - 1550 (Smith and Karpinski 1911: 133)
Greece[13]	12th century: commentaries on Euclid's *Elements* (Wilson 1981)	c. 1400: Ottoman conquest of most Greek-speaking areas
Scandinavia	c. 1275-1300: Valdemar's year-book (Kroman 1974: 120)	c. 1550 (books, manuscripts, records)
Portugal	1415: *Livro da Virtuosa Bemfeitoria* (Barrados de Carvalho 1957: 124-5)	1490 - 1510: travelogues, scientific documents (Barrados de Carvalho 1957: 125)
Russia	Unknown	17th century (reforms of Peter the Great)

In general, Latin and scholarly (particularly mathematical and astronomical) uses of the numerals preceded their vernacular and commercial use by several centuries, with the latter not until the 13th century and not commonly until the late 15th century (Murray 1978: 193-4). Most of the earliest examples of the numerals in any given region are found

[12] Of course, Arabic documents from Spain used the ghubar numerals extensively from the 10th century AD onward; this date refers only to their common use in Christian Spain.
[13] These dates reflect the introduction of Arabic positional numerals (under influence from the Islamic world). The Western numerals were not widely used in Greece until the 18th century and not for administration until Greek independence in 1832.

in mathematical treatises and texts designed specifically to explain the new numerals. Only when the audience for the numerals expanded from monks and scientists do these numerals begin to replace Roman numerals throughout Europe. For instance, in England, the common use of Western numerals was brought about in part by the transmission of double-entry bookkeeping from Italy in the period 1530-1550; this technique requires ciphered-positional numerical notation (Jenkinson 1926: 267). Continental Europeans started to date coins with Western numerals at around the same time, the first being a Swiss coin from 1424, Austria following in 1456, and France, Germany, and the Low Countries in the final quarter of the 15th century (Hill 1915: 94-105).[14] Once records began to be kept and coins minted using the new numerals, their spread to a large segment of the populace was inevitable.

It is interesting to speculate on the possible correlation between the rise in frequency of the Western numerals and the birth of printing in the middle of the 15th century. One of the reasons why it took so long for the Western numerals to become popular is probably the conservatism of medieval churchmen and mathematicians. The rise of literacy after the invention of the printing press, and the consequent expansion of literacy and numeracy to a broader range of people, may have prompted a new willingness on the part of the mercantile class to use the new invention for a variety of purposes, including bookkeeping, inscriptions on coins and seals, foliation, and stichometry. Bibles began being printed using Western numerals in the mid-16th century (Williams 1997). The only study of the relative proportion of Roman and Western numerals in the 15th and early 16th century, that of Barrados de Carvalho (1957) concerning Portuguese texts, confirms that Western numerals began to be used frequently around 1500 or shortly thereafter.

---

[14] A copper coin of Norman Sicily dated to 533 AH (1138 AD) is often given as the earliest positionally-dated coin in Europe, but it is inscribed and dated using the Arabic script and numerals (Hill 1915: 16; Menninger 1969: 439).

The spread of the Western numerals throughout the world, and their eventual replacement of large numbers of indigenous numerical notation systems, could only occur once European countries had become politically powerful. The replacement of numerical notation systems began *en masse* in the 16th century. Within a few decades of Europeans reaching the New World, the Aztec and Maya systems had become obsolete and the Inka *quipu* greatly restricted in scope. This replacement was imposed by colonial powers, who destroyed not only the indigenous numerals of the New World but also many other traditions. At around the same time, the ciphered-additive systems of Eastern Europe and the Caucasus (Cyrillic, Glagolitic, Armenian, and Georgian) began to be replaced by Western numerals or Arabic positional numerals. Finally, the 16th century marked the denouement of the Roman numerals as a system for daily use; coins, documents, and books began to use the Western numerals. By the 18th century, Roman numerals served only archaic and formal functions, ending two millennia of their effective domination in Western Europe. It is remarkable that a system that had not yet been established in its heartland in the 15th century could almost entirely replace not only the Roman numerals but also many other ancient systems in less than three centuries.

The modern era of colonialism brought about the replacement of further systems starting in the 19th century. The Hebrew, Coptic, and Syriac alphabetic numerals all continue to be used for religious and formal purposes, but Western numerals are used in many other contexts. The indigenous numerical notation systems of South and East Asia have not been completely replaced, but they too have been supplanted for many purposes by the Western numerals. While there is no functional reason for the replacement of one ciphered-positional system by another, the dominance of the European nations, coupled with the desire to have a single, universally intelligible symbol system, has made the Western numerals an attractive option. Even in places like Japan and Thailand, which were never under direct political control by a European power, the Western numerals are preferred in most contexts. At the same time, the spread of Western numerals in the 19th and 20th centuries has spawned a host of descendants among North American and African peoples, such as the Iñupiaq, Cherokee, Obɛri Ɔkaimɛ, and Mende systems (ch. 10). Many of these are structurally different from the Western numerals, and are not simply the ordinary system recast with new numeral-signs. It is therefore premature to state that the eventual universality of Western numerals is inevitable.

Summary

We began this chapter with a small set of poorly understood Buddhist inscriptions from India, using hybrid ciphered-additive / multiplicative-additive numerals. We end in the modern era with an enormous variety of local numerical notation systems and two (the Western and Arabic) that spread enormously on the heels of political conquests. With few exceptions, those systems that have survived are ciphered-positional, which surely indicates that such systems are useful, but this certainly does not indicate that they are the inevitable conclusion of a teleological historical process. The common feature of the systems of this family is the set of nine Brāhmī unit-signs, which persist, though greatly altered, in the surviving numerical notation systems of this family, including our own. Only the alphasyllabic systems use distinct signs, the letters of the Indian alphasyllabaries.

It is common practice to end studies of numerical notation with the analysis of Western numerals (e.g. Guitel 1975; Ifrah 1998). Ifrah portrays the spread of Western numerals throughout the world, displacing older systems as it goes, as the inevitable replacement of worse with better systems, in continuous progress from primitive beginnings to the perfection of our own decimal positional system, an achievement which can never be surpassed (1998: 592-3). However, is this an accurate depiction of the history of numerals?

If it were truly the case that technology spreads only through the diffusion of what is functional and the replacement of what is not, we would expect that all systems that are structurally identical should expand with equal rapidity and geographical reach. The South Asian family, with so many decimal ciphered-positional systems surviving and in use, provides a good testing ground for this theory. Because only the Arabic and Western numerals have spread, we can conclude that their diffusion is due mainly to sociopolitical factors. Furthermore, we can look at surviving non-positional systems, such as the Tamil numerals, and see that their geographical distribution is no less widespread than that of many positional ones. How, then, are we to explain why the Tamil system is as widespread as, say, the Khmer, if functionality is of supreme importance? It cannot be done, except by special pleading.

I do not mean to suggest that functionality has nothing to do with the spread of numerical notation systems, especially ones such as the Arabic and Western numerals that have been used extensively for accounting, arithmetic, and higher mathematics. Yet to proclaim the Western numerals' spread as the triumph of functionality and reason over illogic and unwieldiness is to ignore the history of many ciphered-positional systems of this family that have failed to spread – or failed to survive. While the Western numerals comprise one branch of the South Asian family, and a very important one, due to the political might of nations that use them, that is all they are – one branch of many in this family, one of many families to have been used throughout history. In placing the Western numerals in the middle of my study, I choose to emphasize that their apparent triumph is only the present manifestation of one branch of one family. I hope that in doing so, it will become clear that the overwhelming predominance of Western numerals can be seen in a much different light.

Chapter 7: Mesopotamian Systems

The earliest instance of the development of a full-fledged numerical notation system occurred in Mesopotamia. Scholars interested in the diffusion of Babylonian astronomy and mathematics to the Greeks have long studied Mesopotamian numeration (Neugebauer 1957; van der Waerden 1963). Yet to depict the Mesopotamian family of numerical notation systems as an archetypal case for the evolution and diffusion of numerals, or to use its history as the basis for a universal evolutionary pattern, is dangerous (Schmandt-Besserat 1992; Damerow 1996). While Mesopotamian mathematics is extremely important for understanding later Greek developments (and, in turn, modern Western mathematics), Mesopotamian *numeration* is a dead end in the history of numerical notation. Although its history spans three millennia, the Mesopotamian numerals did not spread geographically far beyond their point of origin, and did not survive when placed under pressure from the numerical notation systems of later inhabitants of the region.

The major numerical notation systems of this family are shown in Table 7.1. There are several ways to classify them, depending on which features we want to emphasize. Looking at the numeral-signs alone, the systems divide rather neatly into archaic systems, used prior to 2000 BC and written using curviform symbols made with a round stylus, and cuneiform systems, which developed out of archaic systems and were written using wedge-shaped symbols. Mesopotamian numerals were written almost exclusively on clay tablets by impressing signs onto wet clay using a stylus (though a few stone inscriptions are attested). A second important division is between systems that are primarily decimal and those that are primarily sexagesimal, or base-60. Mesopotamia is the only region of the world where sexagesimal numerical notation is attested (although no system is purely sexagesimal, as we shall see).[1] Finally, comparing the interexponential structure of the systems, we can distinguish between additive systems, which include most of the systems of this family, and positional systems, of which the only true example is the Babylonian sexagesimal positional system.

Table 7.1: Mesopotamian numerical notation systems[2]

System	1	10	60	100	120	600	1000	1200	3600	7200	10000	36000
Archaic systems												
Sexagesimal												
Bisexagesimal												
Bisexagesimal 2												
Proto-Elamite decimal												
Cuneiform systems												
Sumerian												
Assyro-Babylonian												
Mari												
Hittite												
Old Persian												
Babylonian positional												

Proto-cuneiform

Around 3200 BC or perhaps slightly earlier, the antecedent of the later Sumerian script arose at the city of Uruk in southern Mesopotamia, during what is now known as the Uruk IV period.[3] This proto-script, which was probably read in Sumerian, was little more than a set of ideographic signs, lacking any means of expressing phonetic sounds. By the Uruk III period (c. 3000 BC), it had spread from Uruk (the primary Mesopotamian

[1] However, see Price and Pospisil 1966, who claim rather bizarrely that the Kapauku of Papua New Guinea derived their sexagesimal lexical numerals from the comparable Babylonian numerical notation.
[2] This table does not include all the variant systems from the archaic period (for which see below).
[3] Over the past twenty years of research, the chronology of protohistoric Mesopotamia has been shifted back around two centuries; older sources tend to regard the Uruk IV period as representing the early third rather than the late fourth millennium BC.

city at the time) to the north, to Jemdet Nasr, Khafaji, and Tell Uqair. The texts of this period of Mesopotamian history, often called the archaic period, do not represent a true literate tradition but rather a protohistoric system of bookkeeping and administration. In total, about 5,600 clay tablets have been recovered which record this script, known as *proto-cuneiform*. Around 60 of the 1200 proto-cuneiform signs can be assigned numerical or metrological values (Nissen, Damerow, and Englund 1993: 25).

The proto-cuneiform texts are all accounting documents, often written on both sides - the obverse with a series of amounts of commodities, the reverse with a single total. The meanings of the numeral-signs can be deciphered partially by assigning values to the signs and taking the sum of the signs on one side to see if they match with the total on the other. Unfortunately, the assumption that the texts all followed the sexagesimal system – the only one to survive beyond the archaic period – is insufficient to decipher the system completely. Falkenstein, who wrote the first comprehensive description of the Uruk tablets, thought there were separate decimal and fractional numeral systems in addition to the sexagesimal system, and his hypothesis held sway for over 40 years (Falkenstein 1936: 202-214). Unfortunately, for Falkenstein's decipherment to hold true, we must believe that the archaic accountants at Uruk were ridiculously poor scribes and prone to making arithmetical errors and omitting signs. Following this work, Friberg (1978-9, 1984) correctly determined that there was no proto-cuneiform decimal system, but did not come to a full decipherment.

The full decipherment of the proto-cuneiform numerical notation systems did not occur until the 1980s when, through computer-aided mathematical analysis of the entire corpus of texts, Nissen, Damerow, and Englund (1993) established that as many as 15 distinct systems (of which five were particularly common) were used at Uruk.[4] Each system was shown to have been used exclusively for enumerating a specific category of

---

[4] My discussion of the systems below (including their functions) is derived almost entirely from the work of Nissen, Damerow, and Englund (1993: 25-29).

discrete objects or metrological quantity, as indicated by the ideograms found in conjunction with it. The ingenious technique used in their decipherment involves examining the maximum number of times each numeral-sign is repeated to determine the relative values between signs in any given system (as if we were to infer that the Roman numeral V represents 5 by noting that I is repeated four times at most). This technique works because of two properties of cumulative-additive numerical notation systems: the ordering of exponents within each numeral-phrase from highest to lowest and the regular replacement of lower exponents by higher ones wherever possible. A difficulty is that a given numeral-sign can be found in several of the proto-cuneiform systems, but its value often varies from system to system. Thus, ● is equal to 10 D in some systems, but 6 D in others.

A further difficulty is that while we can identify the numerical *ratio* between the values of any two signs within a system, we often cannot identify the specific quantity represented by any one sign, because many systems are used for counting metrological units rather than discrete objects. For systems used for counting discrete objects, it is easy to identify the basic sign for 1 (since, for instance, fractions of humans do not normally occur in texts). I have included absolute numerical values for the four systems that represent discrete quantities in the tables below. For systems that measure area or capacity, we can never ascertain with certainty which sign (if any) has the basic value of one unit. I present the values for these for these metrological systems as ratios, since we can only tell the value of a sign relative to the other signs of the system. Nissen, Damerow, and Englund conclude from this feature of the proto-cuneiform systems that the archaic numeral-signs do not represent abstract numbers at all, but instead are context-dependent numerals that represent concrete quantities within a particular system.

Despite having different numeral-signs and different numerical values, all the proto-cuneiform numerical systems have much in common. All are cumulative-additive, although some individual numeral-signs are formed multiplicatively; e.g. (600) = (60) x ● (10). As in all cumulative-additive systems, the value of the resulting numeral-phrase is read by taking the sum of the individual signs. Groups of identical signs were sometimes sorted into two or three rows for easy reading, but this was not a universal rule, and some tablets contain long strings of signs. Numerals were most often grouped with signs arranged from highest to lowest (although there are some rare exceptions, which may be scribal errors). A single numeral-phrase, together with one or more ideograms, was enclosed in a box in a section of the text.[5]

Sexagesimal systems

The two sexagesimal systems shown in Table 7.2 alternate between factors of 6 and 10, and were the first and easiest to be deciphered because their structure is identical to that of the later Sumerian numerals. The main sexagesimal system (S) is employed in slightly less than half the Uruk texts (Damerow 1996: 292). It was used to count most discrete objects: humans, animals, finished products, tools, and containers, which certainly explains its frequency of use. The subsidiary S' system is used to count a much smaller set of discrete objects, such as jars of some liquids and dead animals.

Table 7.2: Sexagesimal numerals

	36000		3600		600		60		10		1		1/2
Sexagesimal (S)[6]		=10		= 6		=10		=6	●	=10		=2	
Sexagesimal (S')								=6		=10			

Bisexagesimal systems

The two bisexagesimal systems shown in Table 7.3 are so named because an additional factor of 2 is interpolated among the factors of 6 and 10 used in the

---

[5] Note that the proto-cuneiform script was written vertically in columns reading from top to bottom, but I follow Assyriological convention (and that used by Nissen, Damerow, and Englund) in showing the signs rotated 90° counterclockwise and thus read from horizontally left to right. This convention reflects a similar change in the direction of writing cuneiform signs around the middle of the third millennium BC.

[6] The letters in parentheses in this table and the following ones are those assigned to each system by Nissen, Damerow, and Englund (1993) in their research.

sexagesimal systems. While the regular bisexagesimal (B) system is identical to the regular sexagesimal (S) system for all numbers up to 60, it has new signs for the values of 120 (60x2), 1200 (120x10), and 7200 (1200x6). It was used for counting discrete numbers of grain products, cheese, and fresh fish, and is thus the second most common system found in the archaic texts. The function of the identically structured but much less common B* system is unclear but may have indicated discrete quantities of some kind of fish. Both systems appear to have been part of a rationing system, perhaps one for which a number-sign between 60 and 600 may have been useful.

Table 7.3: Bisexagesimal numerals

	7200		1200		120		60		10		1		1/2
Bisexagesimal (B)		=6		=10		=2		=6	●	=10		=2	
Bisexagesimal (B*)				=10		=2		=6		=10			

GAN$_2$ system

The GAN$_2$ system shown in Table 7.4 is used to represent area measures, and is thus the first system I have discussed for which absolute numerical values cannot be assigned. It is peculiar in that, while its signs are the same or similar to those of the common sexagesimal system, the GAN$_2$ signs' values are very different from that of system S. For instance, where ◉ means 36,000 in sexagesimal numerals and is thus 3,600 times greater than ● (10), in the GAN$_2$ system it is only ten times greater (and, moreover, it is six times *less* than ●, which is only 3600 in system S). This similarity probably has something to do with the use of round-ended writing styli in all the proto-cuneiform numerical systems.

Table 7.4:GAN$_2$ numerals

●	=6	◉	=10	●	=3		=6		=10?	

EN system

This rather uncommon notation system, indicated in Table 7.5, is known from only 26 texts. It is unclear what it represents, but it might represent weight measures.

Unusually, all but one of the tablets from Uruk on which the EN system was used were found at a single locus, suggesting that whatever its function, it must have been very restricted in use.

Table 7.5: EN numerals

	=10		=2		=2	

ŠE systems

This relatively common group of numerical systems shown in Table 7.6 was used for various capacity measures of grain. It has caused enormous confusion in the interpretation of archaic texts. While its signs are similar or identical to those of the sexagesimal systems, their order and the ratios between successive signs are quite different. For instance, while the ratio between and is 10 in the sexagesimal, bisexagesimal, and EN systems, it is only 6 in the ŠE subfamily. The regular Š system is used for capacity measures of barley, the Š' system for germinated barley for brewing beer, and the Š* system for barley groats.

Table 7.6: ŠE numerals

System Š		=10		=3		=10		=6		=5			
System Š'						=10		=6		=5		=5	
System Š*				=3		=10		=6		=5			

U$_4$ system

This rather unusual numerical notation system becomes more clearly understood when it is recognized that its function is for recording time and calendrical units. By combining a single ideographic sign with numerical signs for 1 and 10, all the major divisions of the year could be expressed easily.

Table 7.7: U$_4$ numerals

	=10 =12		=3		=10	
10 months / 1 year		1 month		10 days		1 day

The origin of proto-cuneiform numerals

The multiplicity of proto-cuneiform systems and the indeterminacy of their numeral-signs' values are very unusual features that require explanation. While the early date of their use guarantees that their development is a case of local independent invention (quite probably at Uruk) rather than diffusion from an external source, this does not tell us anything about why the numerals took the form they did and whether they have any antecedents.

One of the most popular theories on the origins of Mesopotamian numeration is that they emerged from a system of clay tokens used for accounting in preliterate times.[7] Throughout Mesopotamia and even further abroad, small clay objects of various shapes and sizes have been found in strata dating between 9000 and 2000 BC. Oppenheim (1959) was the first to assign an administrative function to a hollow clay ball, or *bulla*, found at Nuzi inscribed with a brief cuneiform text enumerating 48 animals, and containing within it 48 small stone counters. Amiet (1966) showed that this technique was used much earlier than previously thought (since at least 3000 BC) and that the bullae were 'double documents' through which transfers of goods such as livestock could be conducted while minimizing the risk of fraud or error. A literate official could see the quantity of goods from the inscription on the outside, but if there was any doubt, the bulla could be broken open and the clay or stone tokens inside counted to match them up with the actual quantity received. More recently, Denise Schmandt-Besserat (1984, 1987, 1992) has examined the evidence for Mesopotamian clay tokens, and has concluded that the clay tokens are of much greater antiquity than previously thought and are ancestral to both the proto-cuneiform numerals and the proto-cuneiform script. According to her, the tokens represent a stage of 'concrete counting', fusing quantity (the number of tokens)

[7] Other theories, such as van den Brom's (1969) assertion that the sexagesimal system was based on a prehistoric system of finger-reckoning and Ifrah's (1998: 92-95) claim that it represents a system that originated from the union of two unattested civilizations, one with base-5 and the other base-12 numerals, are entirely hypothetical.

and quality (different shapes representing different commodities), but they do not represent abstract numbers (Schmandt-Besserat 1984: 55). This, she believes, is very similar to the means by which numbers are expressed in the proto-cuneiform numerals, and indeed the theories of Nissen, Damerow, and Englund are partly based on her work. A number of late 4th millennium BC bullae, especially from Susa in modern Iran, are impressed with signs resembling later archaic numerals and contain the correct total of tokens, suggesting that the systems are connected (Nissen, Damerow, and Englund 1993: 127-9). Furthermore, from certain similarities between the three-dimensional tokens and the proto-cuneiform signs, Schmandt-Besserat argues that the tokens developed into writing through the recognition that, if the total of a transaction is written on clay, one need not actually use the clay tokens but need only record their values.

Schmandt-Besserat's conclusions have been received with some skepticism (see especially Lieberman 1980; Zimansky 1993).[8] Firstly, the scope in time and space of the token system is far greater than that of the proto-cuneiform numerals. Lieberman notes, "The examples which she uses range in period from the ninth to the second millennium, and in find-spot from Abydos in Egypt to Iranian Tepe Yahya and Hacilar in Anatolia. The assumption that a single system could have been uniform over such a vast territory and time is untenable" (1980: 352-3). Yet Schmandt-Besserat's study assumes explicitly that this is the case. The geographic origin of the token system also appears to contradict its attribution as the precursor of the proto-cuneiform numerals. In Schmandt-Besserat's study of tokens from the Uruk-Jemdet Nasr period (c. 3000 BC), around two-thirds of the tokens come from Susa in Iran, while only 10% come from the very thoroughly excavated site at Uruk (Lieberman 1980: 353). This suggests that the token system is unlikely to have given rise to numerals and writing at Uruk, where the earliest numerals were

[8] I cannot hope to address her claim that the tokens are ancestral to the proto-cuneiform script, and will restrict myself to the similarities and differences between the token system and the proto-cuneiform numerals.

ound. As well, Schmandt-Besserat has included in her analysis many tokens that were oose rather than contained in sealed bullae, and thus cannot be attributed an accounting unction. Many of these loose tokens are much too large to be considered part of the same accounting system as the accounting-related tokens found within bullae.

Yet the greatest problem with her study, and the one that in my mind refutes it decisively, is the lack of correspondence in shape or structure between the tokens and the proto-cuneiform numeral-signs. Zimansky points out that some of the most common tokens are correlated with proto-cuneiform signs for rare objects such as nails and days of labour, whereas given the accounting function established for the tokens, we would expect livestock, people, and grain to be the most common tokens, as is the case in proto-cuneiform texts (Zimansky 1993: 316). This discrepancy points out a further problem. The archaic numeral systems always place a numeral-phrase in front of an ideographic sign to represent a quantity; "16 + sheep" = "16 sheep", and so on. While the proto-cuneiform numerals partly fuse quantity and quality, because different systems are used for different commodities, they do not do so completely, because one always needs a further sign to indicate exactly what is being counted. With the tokens, however, there is no separation of numerals and the objects being counted; to show 16 sheep, one simply uses 16 tokens for "sheep". Schmandt-Besserat has established some vague correlations between the ideographic signs from Uruk and the shapes of clay tokens, of which some but not all have an accounting function.[9] Yet there is no correspondence between the archaic numeral-signs and the shapes of tokens; thus, to presume that the tokens are ancestral to the numeral-signs is quite fallacious. While the numerical use of tokens antedates the use of the numerals, this does not imply that the tokens gave rise to the numerals. The use of tokens sealed within bullae appears to have been an accounting

---

[9] I do not have the expertise to evaluate Zimansky's (1993: 515) comments regarding the difficulties of correlating the two-dimensional representations on clay tablets with the three-dimensional token system.

technology that predated, but then later coexisted with, the proto-cuneiform numerals. While some early proto-cuneiform numerals are found on clay bullae, this is insufficient evidence that tokens led to numerals.

Another theory explains the origins of the proto-cuneiform numerals far better. Numerical signs resembling the proto-cuneiform ones have been found, not on bullae, but on ordinary clay tablets in late preliterate contexts at Uruk as well as at Jebel Aruda, Susa, and elsewhere (Nissen, Damerow, and Englund 1993: 127-130; see especially Figure 113, 114). These tablets have numerical signs only (*no* ideographic signs), and do not follow the ordinary rule that once a certain number of lower-valued signs have been written, they are to be replaced with a single higher-valued sign. For instance, one tablet from Jebel Aruda contains 3  signs, 22  signs, and at least 5  signs (Nissen, Damerow, and Englund 1993: 130). In any of the later systems, 22 signs would have to be replaced by a smaller number of higher-valued signs. Two important conclusions follow from these tablets. First, because these inscriptions are found in late preliterate contexts and are similar but not identical to the proto-cuneiform numerals, it is probable that they are immediately ancestral to them and date from a period when the system was still being developed. Second, because this early system was used outside the context of the token/bulla system, this further confirms that the token system was not related directly to the development of the proto-cuneiform numerals.

Regarding the structure of the proto-cuneiform numerals, it is not surprising that so many of the systems are structured sexagesimally – or, to be more precise, using multiples of 6 and 10, often alternating. While we do not know the language in which proto-cuneiform numerals were read, the Sumerian lexical numeral system is mainly sexagesimal, and furthermore, 10 is a sub-base in the lexical numerals just as it is in the proto-cuneiform numerals (Powell 1971, 1972a, 1972b). On this basis, Powell (1972b: 172) has correctly discerned that "the presence of a sexagesimal system of notation in the archaic texts from Uruk and Jemdet Nasr constitute [*sic*] the best – indeed irrefutable –

evidence that Sumerian is the language of those texts".[10] It has often been claimed that sexagesimal numeration is a product of the fusion of two prehistoric civilizations, one of base-10 numeration, the other of base-6 (see Thureau-Dangin 1939 for a review). This argument is simply bizarre; a base-10 numeral system would specially denote the numbers 100 and 1000, while a base-6 system would emphasize 36, 216, and 1296 – *not* 600 and 3600. Yet, despite Powell's (1972b) persuasive proof that Sumerian lexical numerals were sexagesimal in the prehistoric period, and his consequent rejection of all such fictitious migrationist explanations, they still continue to enjoy some popularity. For instance, Ifrah (1998: 92-95) claims that sexagesimal numeration in Mesopotamia originated from the union of two unattested civilizations, one with base-5 and the other base-12 lexical numerals. His theory involves hypothetical migrations of *two* unknown civilizations and invokes an entirely unattested form of finger-numerals[11] to produce a system that shows no trace of either base-5 or base-12. The simplest explanation is in this case the correct one: the prehistoric Sumerians, users of sexagesimal numerals that emphasized the number ten, developed written numerals that were largely but not perfectly in accord with their spoken numerals.

A further fact in need of explanation is the multiplicity of proto-cuneiform numeral systems and bases. The origin of this practice is probably not lexical but rather based on Sumerian metrological systems in the Uruk period, for which we unfortunately do not have sufficient evidence. We do, however, have substantial textual evidence for the metrological systems of the Early Dynastic and later periods. The ratios between various signs in the proto-cuneiform numeral systems dealing with measures of capacity, area, and weight are similar to the ratios found in Sumerian metrological systems of the

---

[10] It is possible, though unlikely, that the language of the system's users was not Sumerian, but rather an unattested language (possibly related to Sumerian) that also had sexagesimal lexical numerals.

[11] A similar hypothetical finger-reckoning argument is presented by van den Brom (1969), which is somewhat similar to Ifrah's theory (though Ifrah does not appear to be aware of it).

Early Dynastic period. This supports the contention that the odd ratios of some of the older systems are due to unattested metrological systems that continued into better-documented periods, and further confirms the decipherment of the systems presented earlier.

All scholars are agreed on a number of crucial points regarding the origin of proto-cuneiform numerals. They were a local Mesopotamian development in the late 4th millennium BC (possibly as early as 3500 BC, but more likely closer to 3200 BC). One of the early uses for these numerals was to imprint values on clay bullae that were used in conjunction with tokens for accounting purposes. A large number of distinct systems were developed (at least 15, and probably more) and all served strictly economic and administrative functions. Yet despite my agreement on these issues, I do not see enough evidence to conclude that the tokens were ancestral to proto-cuneiform numerals. Rather, the increasing administrative demands that developed with the rise of the Uruk city-state in the late 4th millennium BC created a new need for record-keeping, metrology, and accounting, of which the numerals and the clay bullae are two distinct consequences.

Cognitive consequences of proto-cuneiform numeration

The analysis of the proto-cuneiform numerals has also led researchers to speculate on the possible cognitive correlates of the use of multiple numerical notation systems. Damerow (1996) has argued that the material record from the archaic period in Mesopotamia is a direct reflection of the numerical abilities of Mesopotamians.[12] Furthermore, he contends that the tokens and proto-cuneiform numerals can be used to reconstruct a universal stage of concrete numeracy that precedes the modern abstract number concept. In this respect, his argument is similar to that of Hallpike (1979), who

---

[12] Similar arguments are raised by Schmandt-Besserat (1992) regarding the system of clay tokens, but I will not address these in detail here, given my rejection of her interpretations of the token system above.

applies the insights of Piaget, Vygotsky, and others from developmental psychology to draw a parallel between individual cognitive development and the evolution of thought in societies.[13] I am unconvinced that the use of multiple proto-cuneiform systems at Uruk tells us much about the cognitive capacities of ancient Mesopotamians.

Damerow claims that the peoples of early Mesopotamia could not conceive of abstract numbers, but rather were only capable of concrete counting (Damerow 1996: 275-297). As indicated above, multiple proto-cuneiform numerical notation systems were used for representing different categories of object, and a single sign could have different relative values in different systems. From this, he argues that Mesopotamian scribes could conceive of "8 sheep" or "8 jars of oil" but not simply "8" as an abstract concept. Taken to its logical conclusion, this would imply that users of the proto-cuneiform numerals could see nothing in common between 8 sheep and 8 jars of oil. I cannot see how this can be the case; if so, it would be impossible to make the connection between 8 sheep and 8 marks on a clay tablet, and numeration would be impossible. Numeration in cumulative-additive systems is a matter of one-to-one correspondence – that is, associating one set of objects with another by pairing off individual objects in each set.

Furthermore, the postulate that these context-dependent numerals represent a stage of "archaic arithmetic" in a unilinear and universal scheme for the evolution of the number concept cannot be sustained (Damerow 1996: 296). Of all the independently invented numerical notation systems I have studied, only the early Mesopotamian case (and possibly the Inka *quipu*, which I will discuss in Chapter 10) use different means of representation for different types of object. Regardless of the cognitive consequences of the Mesopotamian numerals, there can be no "stage" of concrete numeration when nothing of the sort can be found in Shang, Predynastic Egyptian, or Zapotec inscriptions. This is not to say that there are no cognitive consequences to the use of a dozen or more

---

[13] It does not appear that Damerow is familiar with Hallpike's work.

numerical notation systems, but whatever they are, they will not be universally applicable to every society.

Finally, I reject Damerow's conclusion because we have little evidence for the social contexts of the use of numerals. We have no idea how many of these systems would have been known to any individual official, and no evidence from the archaic period as to how numerals were manipulated and used arithmetically. We simply have values and totals, which do not tell us very much about how people were actually thinking about number.[14] Even if individuals used many systems, this does not prove concreteness of thought. One thing we *do* know (if we presume that the Uruk scribes were speaking Sumerian) is that, in contrast to the proto-cuneiform numerical notation, there was a single perfectly ordinary set of Sumerian lexical numerals (Powell 1971).[15] There are sensible reasons why someone capable of abstract thought would use multiple systems of numerical notation, such as to prevent confusion as to the type of thing being counted. There is no qualitative difference between the Uruk systems and the modern use of Roman numerals to distinguish the foreword of a book from its main text, or the use of hexadecimal numerals for computing purposes. Ironically, one of the principles behind the 'new mathematics' movement in North America in the 1960s was the claim that teaching students to calculate using numerical systems of different bases would *improve* their understanding of abstract number concepts. As indicated above, the rationale behind the origin of many proto-cuneiform systems was to produce a good fit between metrological and numerical systems. If so, the Uruk scribes probably had an

---

[14] Liverani's (1983) intriguing conclusion that a fragmentary Uruk IV-period clay tablet indented with holes may have served as a counting board has not been confirmed and must remain tentative unless further finds are made.
[15] Schmandt-Besserat (1984, 1992) has made much of the parallel between the many proto-cuneiform numerical notation systems and the use of 'numerical classifiers' in Japanese, the Mayan languages, and others, where the set of numerals is modified depending on the class of object being counted. Notably, numeral classifiers are *not* a feature of Sumerian, and this theory, if taken to its logical conclusion, implies that the modern Japanese do not have a concept of abstract number.

abstract number concept, but realized that abstract written numerals were not the most efficient solution to the problems they were facing. The theoretical importance attributed to the proto-cuneiform numerals as evidence of an evolutionary stage of cognition is entirely unwarranted.

Convergence and decline

While they are an interesting early example of numerical notation, the proto-cuneiform numerals did not diffuse extensively or last for an extended period. There are no significant resemblances between the proto-cuneiform numerals and the Egyptian hieroglyphic numerals (ch. 2), which may precede the proto-cuneiform systems in any case. The only systems that were obviously borrowed from Mesopotamia at this time are the proto-Elamite systems used from about 3000 BC at the site of Susa and elsewhere in modern Iran. I will discuss these systems below.

The start of the Early Dynastic period in Mesopotamian history marked a turning point in the history of its numerals. Beginning around 2900 BC, there was a marked decline in the frequency of almost all the proto-cuneiform numerical systems, while the sexagesimal system rapidly assumed the functions of the other systems. While the system for measuring area (GAN_2) continued to be used as late as the Fara period (c. 2500 BC), it was clearly in decline and considered archaic by that point (Nissen, Englund, and Damerow 1993: 137-38). While each metrological system had its own numerical notation system in the archaic period, eventually officials decided it was better to express all numbers, regardless of function, using a single notation.

There are three plausible explanations for this convergence, which are not mutually exclusive of one another. The simplest is that the use of so many systems in so many different functions was cumbersome for administration, potentially confusing, and open to abuse. This may simply be a modern prejudice attributable to the Western use of only one set of numerals. While 200 years is a short time in the context of world history, it is a long time for a truly inefficient set of systems to persist. In addition, while the archaic texts were used at only a very few locales (mainly at Uruk), the later numerals were used throughout Mesopotamia. The Early Dynastic period was marked by relative political stability and multiple alliances among Sumerian city-states. If the Early Dynastic period marks the first era when Mesopotamian numerals were employed for long-distance communication, the use of a single system to facilitate communication among a larger group of individuals would be advantageous. The sexagesimal numerical notation system is similar to the Sumerian lexical numerals, which were used throughout Mesopotamia at this time (Powell 1971). Finally, changes in Sumerian metrological systems may have reduced the usefulness of the proto-cuneiform systems by eliminating the fit between metrology and numeration.

Proto-Elamite

Around 3100 BC, a ideographic writing system developed in southern and western Iran, the region that would be known as Elam in later Mesopotamian sources. This script, now known as "proto-Elamite", is attested in over 1500 texts, mainly from the major urban centre of the region, Susa; most date from the Susa III period around 3000 BC. A few other proto-Elamite texts have been found at Tepe Yahya and elsewhere in modern Iran. It is a linear script, read from right to left and in lines proceeding from top to bottom. Many mysteries remain with regard to this script, because the language it was intended to represent cannot be identified. One fruitful area of study has been the analysis of proto-Elamite numerals. While the signs of the proto-Elamite script are entirely different from those of early Mesopotamia, the proto-Elamite numerals are very similar to the proto-cuneiform systems and are clearly descended from them.

Yet, as with the proto-cuneiform numerals, confusion over the nature and number of proto-Elamite numeral systems has delayed their correct decipherment until recently.

Brice (1962-3) provides a useful summary of several early twentieth century efforts to decipher the proto-Elamite numerals, all of which assume a single decimal and cumulative-additive numerical notation system. Ifrah (1998), whose interpretation was largely developed in the 1970s and early 1980s, believed there to be two systems, one primarily decimal, the other of a mixed decimal-sexagesimal structure. These early efforts assumed that the relative values of any two signs are fixed.

An adequate decipherment of the proto-Elamite numerals has been achieved recently through the mathematical analysis of the corpus of proto-Elamite texts by Robert Englund and Peter Damerow (Damerow and Englund 1989; Englund 1996). Damerow and Englund realised that, as with the proto-cuneiform numerals, not only were there multiple proto-Elamite numerical notation systems, but the relative values of individual numeral-signs vary from system to system. There are five major proto-Elamite systems: three for counting discrete objects, another (with three variants) for capacity measurements, and another for area measurements (Englund 1996: 162).

The proto-Elamite numerical notation systems for counting discrete objects are shown in Table 7.8 (Englund 1996: 162; cf. Potts 1999: 78).[16]

Table 7.8: Proto-Elamite numerals (discrete objects)

	1	10	60	100	120	600	1000	1200	3600	10000	Function
Sexagesimal											Inanimate objects
Bisexagesimal											Grain products
Decimal											Animate objects

Obviously, the three systems are identical for 1 and 10, and the sexagesimal and bisexagesimal systems are further similar for 60. The sexagesimal system, like the proto-cuneiform sexagesimal numerals, is not a pure base-60 system; instead, each successive

[16] As with the proto-cuneiform numerals, I have represented the numerals as they would be read horizontally (following Assyriological convention; cf. Damerow and Englund 1989) rather than vertically (cf. Englund 1996).

number alternates by factors of 10 and 6. In the bisexagesimal system, the value 120 comes after 60 (a factor of 2). The decimal system is purely base-10 (making it unique among Mesopotamian numerical notation systems in *not* having a mixed or irregular base). It has no correlate in the proto-cuneiform numerals (though its signs are probably derived from those of the bisexagesimal system, which is of course attested in proto-cuneiform).

The main signs of the systems for measuring capacity and area are shown in Table 7.9. Because they are not used for discrete objects, they are represented in terms of the ratios between values, not as discrete numerical values. These two systems are very similar (though not identical) to the ŠE and GAN_2 proto-cuneiform systems, so, following Damerow and Englund, I have used these labels.

Table 7.9: Proto-Elamite metrological numerals

Capacity (ŠE)		=6		=10		=3		=10		=6		=5		=2	
Area (GAN_2)								=10?		=3		=6			

The striking resemblances between the proto-cuneiform and proto-Elamite numerals make it certain that the latter were modelled on the former (Potts 1999: 76-77). In fact, while the respective scripts are entirely dissimilar, it is a matter of personal preference whether we regard the proto-cuneiform and proto-Elamite numerals as distinct sets of systems or rather as two regional variants of a single tradition. Because the first texts from Uruk date to the 33rd century BC, while those found at Susa date to the late 31st century BC, most scholars agree that proto-Elamite ones cannot have been ancestral to those at Uruk, and that they must have diffused from west to east in the context of interregional trade. Given the importance of the Uruk city-state in the late 4th millennium BC, it is unsurprising that the numerals would spread to Susa, the other major polity at that time. The main difference between the two sets of numerical notation systems is the existence of a decimal system for counting discrete quantities of animals and humans in proto-Elamite while no such system exists in proto-cuneiform. While we do not know why this is the case, it may be that the language of the writers of the proto-Elamite texts had decimal lexical numerals, whereas we know that Sumerian numerals are primarily sexagesimal.

The proto-Elamite numerals did not spread beyond Susa and a few other sites in modern Iran. Brice's (1963) tentative identification of similarities between the proto-Elamite and Linear A (Minoan) numerals cannot be taken seriously as indicative of a historical connection. While the proto-Elamite decimal cumulative-additive system is structurally identical to the Minoan, the absence of formal resemblances in the numeral-signs and the enormous geographical and temporal distance between the two civilizations make such a hypothesis highly improbable. The proto-Elamite numerals ceased to be used around 2900 BC, following the decline of Susa as a major urban polity in the early part of the 3rd millennium BC and the subsequent rise of the various Mesopotamian city-states. The numerals in the Old Elamite texts, which are roughly contemporaneous with the Old Akkadian texts in Mesopotamia, are clearly derived from later Mesopotamian systems rather than from Proto-Elamite (Potts 1999: 79). The proto-Elamite numerals are best seen as brief florescence within a single city-state, rather than part of a longer tradition.

Sumerian

The only system among the multitude of proto-cuneiform systems to survive into the Early Dynastic period (c. 2900 to 2350 BC) was the sexagesimal (or more accurately, the decimal-sexagesimal) system. While it was originally used only for counting certain discrete objects, in the Early Dynastic period it was used for all numerical functions, as the older metrological systems were abandoned. At the beginning of the Early Dynastic, significant changes were taking place in the script of the region. The older ideographic and curviform proto-cuneiform symbol system slowly transformed between 2900 and 2500 BC into a writing system that used wedge-shaped cuneiform signs and expressed phonetic as well as conceptual information. From this, we can tell that Sumerian was the language in which the script was read. Yet, despite the radical alterations that the script was undergoing, the numeral-signs remained essentially identical to the archaic sexagesimal ones. One important change occurred around the 27th century BC, when the numerals, like the entire script, underwent a 90-degree rotation, so that they were written and read horizontally from left to right rather than vertically from top to bottom (Ifrah 1998: 84). These numerals are shown in Table 7.10 (Nissen, Damerow, and Englund 1993: 28).

Table 7.10: Sumerian sexagesimal numerals

	1	10	60	600	3600	36000
Vertical						
Horizontal						

These six numeral-signs were combined to make a cumulative-additive numerical notation system. Normally, groups of four or more signs were arranged in two rows to facilitate rapid reading. Thus, 14254 might be written as follows:

14254 = 

(3x3600)+ (5x600) + (7 x 60) +(3x10)+(4x1)

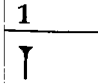
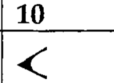
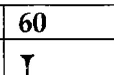
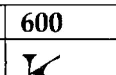
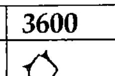
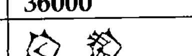
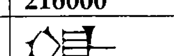
Because this system has signs for both 60 and 3600 (=60²), it obviously has a sexagesimal component. In a purely sexagesimal cumulative-additive system, one would need to repeat each sign up to 59 times, which is clearly not a practical option, but the Sumerian system is not purely sexagesimal, as it has special signs for 10, 60x10 (600), and 3600x10 (36,000). The latter two signs are multiplicative combinations of the small circle for 10 with the sexagesimal signs for 60 and 3600. This decimal sub-base is similar, but not identical, to the use of the sub-base of 5 in the Roman numerals. While the figures of he Roman sub-base (V, L, D) could only occur once in any numeral-phrase because 5, 50, nd 500 are half of 10, 100, and 1000, respectively, the decimal signs in the archaic Mesopotamian numerals could be repeated up to five times, as necessary. The sign for 60 s simply a large version of the sign for 1, just as the sign for 3600 is a large version of the sign for 10. Because the "big 1" is 60 times greater than the regular 1, but the "big 10" (3600) is *360* times greater than its counterpart, I cannot agree with Lieberman's (1980: 343) suggestion that these signs represent the use of 'size-value', which then evolved into 'place-value' over time. This feature is not particularly relevant to the structure of the system, but simply derives from the fact that two styli, one twice as large as the other, were used to impress numerical signs on clay tablets (Powell 1972a: 11-12).

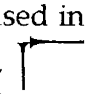
A more notable structural feature of the Sumerian numerals is the first evidence, in the Early Dynastic period, for the use of subtractive notation to express certain numbers, especially those that end in 8 or 9 in the Western numerals. Thus, instead of writing 19 as one sign for 10 plus nine signs for 1, it could be written as , or 20 – 1. The sign is a Sumerian ideogram, LAL. A sign or signs placed inside the LAL sign indicated an amount to be subtracted from the signs preceding it. This technique was used at Fara perhaps as early as 2650 BC and is a regular feature of Mesopotamian numerals from 2500 BC onward (Ifrah 1998: 89). There is no evidence of subtractive lexical numerals in Sumerian as there are in Latin *duodeviginti* and *undeviginti*; the Sumerian words for 18 and 19 are etymologically '10+5+3' and '10+5+4', respectively (Powell 1971: 47). Rather, it seems that this innovation had its origin strictly in numerical notation and the desire to express numbers more concisely.

As in the archaic period, Early Dynastic numerals are found overwhelmingly in documents that served an economic or administrative function. The Early Dynastic period also provides us with the first evidence for the use of Sumerian numerals for arithmetical purposes. In the archaic period, there was no indication how calculations were being done (though of course calculations must have been made). In contrast, ubstantial evidence from Fara (ancient Šuruppak) indicates that some numerals written n clay recorded intermediate steps in calculation and thus were arithmetical aids Høyrup 1982). Similarly, Sumerian 'tables of squares', geometrical and arithmetical xercises, and other arithmetical aids have all been found at Fara (Powell 1976). Nevertheless, the use of numerals for representation, especially in administrative ontexts, exceeds by several orders of magnitude the frequency of their use for omputation. There is nothing indicating the direct use of Sumerian numerals for omputation (by lining up columns, etc.) of the sort common in Greek and Western arithmetic, which is unsurprising given that the Sumerian system is cumulative-additive. Damerow (1996: 236-7) laments the fact that, despite the wealth of Early Dynastic economic records, we have no idea how multiplication was performed; he suggests that it must have been through a non-permanent means, such as counting-boards, finger-reckoning, or mental calculation. Nevertheless, Ifrah's (1998: 123-127) extensive speculations on such hypothetical techniques are unhelpful given the paucity of relevant evidence at present.

By 2500 BC, the transition from the older Sumerian script to cuneiform signs had been completed, except for the numerals. Beginning in the Presargonic period, the older curviform numerals began to be replaced with a set of numerals that used cuneiform numeral-signs, while remaining virtually unchanged in terms of structure (Powell 1972a: 13). This had the advantage of requiring only one stylus for all writing, whether lexical or numerical. While this trend appears to have been initiated by the Sumerians themselves, it was hastened considerably, starting around 2350 BC, by the rise of Akkadian hegemony over Mesopotamia. These new numeral-signs are shown in Table 7.11 (Powell 1971: 244).

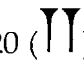
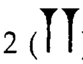
Table 7.11: Sumerian cuneiform numerals

1	10	60	600	3600	36000	216000
						

The signs for 1 and 60, which had previously been semicircular and horizontal, became vertical wedges. I do not know on what basis Thureau-Dangin (1939: 106) asserts that this was derived from an earlier tradition of using vertical strokes for units. The sign for 1 and the sign for 60 were not identical in the earliest cuneiform numerals. Instead, the sign for 60 was written as a "big 1" just as it had been in the curviform numerals, but because the two signs were made with the same stylus, the size difference was always minimal, and soon the two signs became identical (Powell 1972a: 13). This feature certainly does *not* mean that the system is a place-value or positional one, although it may have played a role in the invention of the later sexagesimal positional system (Powell 1972a: 13-14). The old round sign for 10 was replaced by a *Winkelhaken* or corner wedge, made by impressing the stylus perpendicular to the clay tablet, while the large round sign for 3600 was represented visually by four (or occasionally five) wedges placed in a rough circle. In other respects - the writing of 600 and 36000 as 60x10 and 3600x10, and of course the basic cumulative-additive structure - the cuneiform numerals were identical to the curviform ones. The phrase used for 216,000, which is not to my knowledge attested in the earlier numerals, is a combination of the sign for 3600 and the ideogram GAL "big", and is quite rare (Powell 1972a: 7). Powell also describes an even more complex phrase for 12,960,000 (216,000 x 60), šargal šunutaga, 'big everything which hand cannot touch'. For such lexical phrases, we need to ask at what point a phrase ceases to become part of a numerical notation system. The subtractive ideogram LAL is used in this system, as in the archaic one, but it is depicted using two cuneiform wedges ().

The replacement of the curviform by cuneiform numerals was by no means an immediate one, and was not complete until around 2050 BC (Powell 1972a: 13). The older system, while it required additional styli to write numerals, and though its numeral-signs generally took up more space, had the advantage of standing out more clearly in a text of cuneiform characters, thus making totalling easier (Powell 1972a: 12).

Additionally, a norm developed by which the two sets of numerals could be used side by side to indicate different functions. The means by which these categories were determined remains unclear, however. Ifrah (1998: 90), following Lambert, indicates that cuneiform numerals were used for those of high social standing, while the older numerals were used for slaves or common people. Lieberman (1980: 344-5) believes that the older numerals indicate quantities directly counted, possibly using clay counters, while the cuneiform numerals were used for quantities that were determined indirectly, such as for objects not actually present to be counted. Damerow (1996: 238) notes that some Early Dynastic economic texts from Girsu use the older numerals for amounts of grain and the cuneiform numerals for amounts of animals, and hypothesizes that this may have been done to avoid confusing the two different categories when taking sums. While we do not know which, if any, of these theories is correct, it suggests a parallel both with the proto-cuneiform systems and the modern use of Roman and Western numerals side by side, and is further evidence of the functionality rather than the inefficiency of the use of multiple numerical notation systems simultaneously. Nevertheless, the round numerals were completely abandoned by the Ur III period (when Sumerian rulers gained control of Mesopotamia), and are not attested later than 2050 BC (Powell 1972a: 13).

The Akkadian conquest, while probably the most important political event of 3rd millennium BC Mesopotamia, does not appear to have had much effect on numeration, and was not the cause of the shift from curviform to cuneiform numerals. The Akkadian kings and officials (c. 2350 - 2150 BC) were content to use the cuneiform and even the archaic numerals for most of the same purposes as they had been used in the Early Dynastic period. In fact, more change in the numerals is visible in the Neo-Sumerian Ur III period (2150 to 2000 BC), during which the archaic numerals disappeared entirely. One slight modification that was tried in some Akkadian texts was to write multiples of 60 using units followed by the Akkadian lexical numeral for 60, *šu-ši* (𒋗𒅆) using multiplicative notation (Labat 1952: 244-247). Thus, instead of writing 120 as , it would be written as . This is obviously a much more cumbersome representation, and probably was used in part to distinguish 120 () from 2 (). For the higher decades - 70, 80, and 90 - the regular Sumerian forms were always used by the Akkadians (, , and , respectively). In any case, many Akkadian inscriptions where *šu-ši* could be used are written in the ordinary Sumerian fashion, so I do not agree with Ifrah (1998: 142) that this technique represents a distinct Sumero-Akkadian numerical notation system.

The Sumerian cuneiform system is ancestral to all the later systems of Mesopotamia. The Semitic tradition of cuneiform decimal numerical notation systems, including first the Eblaite system and later the Assyro-Babylonian common system, are derived from a Sumerian ancestor. The decimal structure of these systems reflected the Semitic languages of its users, which had decimal lexical numerals. While Thureau-Dangin (1939: 107) believed this tradition to have been developed in the Old Akkadian period (starting c. 2350 BC), it is now clear from the library at Ebla that it developed as early as 2500 BC (Pettinato 1981). The sexagesimal, cumulative-positional system, used in Babylonian mathematics and astronomy, was also modelled on the Sumerian cuneiform system. It may have arisen in the Ur III period, and was used by the 20th century BC at the very latest (cf. Powell 1976, Whiting 1984). While this system preserved the old Sumerian base of 60, it was very different in structure, being the earliest positional system ever developed.

The Sumerian cuneiform system continued to be used for most purposes until the Old Babylonian period (c. 2000 – 1595 BC), at which time it began to be replaced by its two descendants. The Assyro-Babylonian decimal system began to be used for most administrative, commercial, and literary functions, while the sexagesimal positional system was used for mathematics and astronomy – once again perpetuating the tradition of using multiple numerical notation systems for multiple purposes. Several Old

Babylonian tablets provide translations from the old Sumerian additive numerals to the new positional system (Nissen, Damerow, and Englund 1993: 146-7), indicating either a need to learn the new positional system or, alternately, that the older cuneiform system was already being forgotten. By the 15th century BC, it had disappeared from regular use. However, a peculiar vestige of the Sumerian system persisted well into the 1st millennium BC in certain late inscriptions, such as those indicating the sizes of buildings (De Odorico 1995: 4). Such a relict is found on an 8th century BC text dating to the reign of the Assyrian king Sargon II, where the dimensions of the fortress at Khorsabad – 16,280 cubits – are written in a modified form of the Sumerian fashion, using cumulative-additive sexagesimal numerals (Ifrah 1998: 141). It is unclear what the significance of this relict might be.

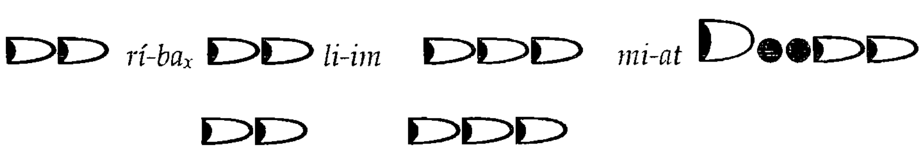
Eblaite

The inhabitants of the city-state of Ebla (in the western part of modern Syria) spoke a West Semitic language but were strongly influenced by Sumerian culture. A great library of thousands of Eblaite cuneiform texts dating certainly to the period prior to 2350 BC (the Akkadian conquest) and possibly as early as 2500 BC, provide us with ample evidence regarding the system of numeration used by the Eblaites (Pettinato 1981). This system is almost identical to that used by the Babylonians some centuries later, and reflects the shared Semitic language and culture of these two groups in contrast to those of the Sumerians. As indicated in Table 7.12, the numerical notation system used at Ebla consisted of two sets of numeral-signs for numbers below 100, one curviform and the other cuneiform (corresponding with the Sumerian archaic and cuneiform systems, respectively), but only one set of signs for the exponents above 100 (Pettinato 1981: 183-184).

Table 7.12: Eblaite numerals

1	10	60	100	1000	10000	100,000
			*mi-at*	*li-im*	*rí-ba*$_x$	*ma-i-at* OR *ma-i-hu*
						

The Eblaite system is cumulative-additive for values less than 100, and multiplicative-additive above that point. The signs for 1, 10, and 60 are ideographic signs identical to those used in the two Sumerian sets of numerals. The two sets of numeral-signs served quite separate functions: the curviform numerals were used for basic enumeration and counting discrete objects, while the cuneiform numerals were used only for capacity measures such as the *mina* and *gubar*, as well as for regnal years of kings (Pettinato 1981: 183-4). The sign for 60 was used to express the tens values in numbers between 60 and 99; its presence, clearly derived from its Sumerian ancestor, is the major irregularity in an otherwise perfectly decimal system. The 'signs' for numbers above 100 are in fact the Eblaite lexical numerals and combined multiplicatively with the unit-signs as necessary. Thus, 24682 could be written as follows:



2 10,000 4 1000 6 100 60 10 10 1 1

Because it is decimal and multiplicative-additive above 100, this system required only one ideographic sign (the crescent or vertical wedge) for the higher exponents; however, the repetition of intraexponential signs for the units, coupled with the use of complex two and three-syllable exponent-signs, meant that numerals were fairly long and cumbersome. To reduce this length, two features were often used.[17] Firstly, just as in the Sumerian system, subtractive numerals were sometimes used for certain numbers to eliminate the need to write 7, 8, or 9 unit-signs by placing the subtrahend after the syllable *lal* or *lá*. Subtractive numeral-phrases were not used consistently throughout the

[17] Strangely enough, Pettinato (1981) does not mention either of these two features, though they are obvious from the many inscriptions that he transcribes in detail.

system. Secondly, the words *mi-at* for 100 and *li-im* for 1000 were often shortened to single syllables *mi* and *li* respectively. Thus, in one text, 7879 (expressing a number of *gubar* measures of barley) is written in cuneiform numerals as 7 *li* 8 *mi* 60 10 10 *lá*-1 (7 x 1000 + 8 x 100 + 60 + 10 + 10 - 1) (Pettinato 1981: 134).

The majority of the Eblaite texts served economic or metrological functions. It is not clear whether the Eblaite numerical notation was ancestral to the later Assyro-Babylonian system or whether the latter developed out of the Sumerian cuneiform system in parallel to the Eblaite system. Because the two systems are very similar in structure (even including their common use of multiplicative structuring above 100 with abbreviated lexical numerals), the possibility that the earlier Eblaite notation was borrowed by the Babylonians seems to be a good working hypothesis. The Eblaite system did not persist past about 2300 BC, after which point Ebla came under Akkadian, and later Amorite, control.

Assyro-Babylonian common

Over the past century, enormous attention has been paid to the Babylonian positional numerals - the cumulative-positional, base-60 system used for astronomy and mathematics. No doubt, this is because historians of mathematics are interested in the origins of our base-60 units of time and the division of the circle, which are derived from this system. The far more common decimal and additive numerals used for most economic, monumental, and literary purposes throughout Mesopotamia are almost forgotten in this analysis of what I call the "Assyro-Babylonian common" system. Though it may have had as its antecedent the Eblaite system, which was used for a couple of centuries in the Presargonic period, it came to assume a dominant position in the Old Babylonian period (starting c. 2000 BC), a position it would maintain for over

1500 years. The numeral-signs of this system are shown in Table 7.13 (De Odorico 1995: 4).

Table 7.13: Assyro-Babylonian common numerals (De Odorico 1995: 4)

1	10	60	100	1000
𒁹	𒌋	𒁹 or 𒋗𒅆	𒈨	𒌋𒈨

This system is cumulative-additive below 100, multiplicative-additive above 100, and is always written from left to right. For the most part, it is purely decimal. The units were expressed cumulatively, except that 9 could be written using three overlaid vertical strokes (𒑆) as an alternative to writing it with nine strokes (𒐎) (De Odorico 1995: 4n). The tens values were most often expressed decimally using one through nine *Winkelhaken* corner wedges for 10. The vertical wedge for 60 is identical to that for 1, but unlike the Sumerian system, 𒁹 was not normally used to represent 60 alone, which would have created ambiguity. It was used in combination with signs for 10 and 1 to write numbers from 70 to 99. Additionally, as in the Akkadian variety of the Sumerian cuneiform system, the lexical numeral *šu-ši* (𒋗𒅆) could be used to indicate 60 or multiples thereof; however, this phonetic form was not used to write 70, 80, or 90. Therefore, there were as many as three different numeral-signs for 60, one decimal (𒐑) and two sexagesimal (𒋗𒅆 or 𒁹), although the vertical wedge could not represent 60 alone. In general, decimal forms appear to be more common than sexagesimal ones for most purposes. Above 100, the multiplicative principle was used quite freely and could be combined in various ways to express very high numbers. The sign for 100 was the syllabic sign ME, an abbreviation of the Babylonian word for 100, *me'at*, while that for 1000 was no more than a multiplicative combination of the signs for 10 and 100. For instance, a scribe from the period of Sargon II wrote 305,412 as 𒐈𒈨𒐊𒌋𒈨𒐉𒈨𒌋𒈫, or ((3 x 100 + 5) x 1000) + (4 x 100) + 10 + 2 (Ifrah 1998: 139). In theory, this system could be extended as far as one wished by juxtaposing signs for 100 and 1000 repeatedly, even though there was no sign for zero. Furthermore, unlike the Sumerian system, in which the signs for 1 and 60 were identical, this system presented no ambiguities to the reader.

Although Thureau-Dangin (1939) and Ifrah (1998) both consider this system to be "Akkadian", this does not seem particularly appropriate, given that it was used only rarely during the period of Akkadian control of Mesopotamia and only began to predominate during the Old Babylonian period. They are certainly correct, however, in attributing its origin to the increased power of Semitic peoples in Mesopotamia in the middle of the 3rd millennium BC: Akkadians, to be sure, but also Eblaites, Babylonians, and others. This system was first used extensively starting around 2000 BC. Its structure reflects the decimal lexical numerals of the Semitic languages rather than Sumerian lexical numerals, although the continued use of a special sign for 60 gives testament to its descent from the Sumerian numerals.

All of the administrative, commercial, literary, and religious texts of the Babylonians and the Assyrians were written using this set of numerals. It is certainly problematic that so much attention is paid to the positional numerals, while this system, which was used over a longer period and had a much greater number of users, is ignored. Many fruitful lines of research into the functions of various systems remain to be investigated. Given that both the positional system and the much older Sumerian system coexisted with the common system for centuries, the question arises as to who would have known and used which system(s), and in what contexts each system would have been used.

Perhaps the greatest significance of the Assyro-Babylonian common system is the large number of descendant systems it produced (as compared to the positional system, which has only one direct descendant). Earliest among these is the system used at the city-state of Mari around 1800 BC, which blends features of this system and the Babylonian positional system. In the middle of the 2nd millennium BC, both the Ugaritic and Hittite cuneiform scripts began using numerals based on the Assyro-Babylonian ones, which is unsurprising, given the importance of Mesopotamian trade with these regions. The Old Persian cuneiform numerical notation system, developed in the 6th century BC (by which time Mesopotamia was under Persian rule), also derived from the Assyro-Babylonian system rather than any of the numerous other systems in use in the region by that time. Finally, and perhaps most curiously, as I argued in Chapter 3, the structural similarities between the Mesopotamian and Levantine families of numerical notation systems are sufficiently strong to postulate that the earliest Levantine systems (Phoenician and Aramaic) were developed as a blend of Egyptian hieroglyphic (or perhaps Hittite) and Assyro-Babylonian influences. The Levantine systems are all decimal, cumulative-additive systems and are all multiplicative-additive above 100. Although they are linear scripts rather than cuneiform ones, these structural similarities, coupled with evidence of considerable cultural contact between the two regions between 900 and 700 BC, strongly suggest a historical connection.

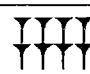
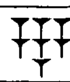
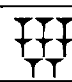
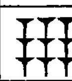
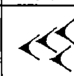
The Assyro-Babylonian additive system flourished despite enormous political changes. It was the system used for administration and commerce by both the Babylonians and the Assyrians until the Persian conquest of Babylon in 539 BC. Afterwards, it began to be supplanted by the Old Persian cuneiform system and, more importantly, by the Aramaic system that then became the principal administrative and commercial system of the region. Of course, both these systems were indebted greatly to their Assyro-Babylonian ancestor. It is unclear when the Assyro-Babylonian system disappeared entirely, but it was used at least to a limited extent throughout the period of Achaemenid rule (539 - 332 BC), and perhaps somewhat later.

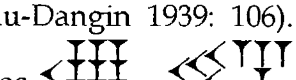
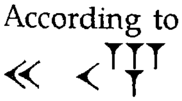
Babylonian positional

The Babylonian positional numeral system is assigned such great importance by many historians of mathematics that one could easily get the impression that it was the only form of Mesopotamian numeration worthy of note. Despite Neugebauer's (1957: 17) warning that the positional numerals are a relatively minor part of the body of Babylonian numerals, these sexagesimal positional numerals, used for mathematics, have been assigned priority over much more widespread systems (Sumerian and Assyro-Babylonian). In fact, positional numerals were used in only a limited set of mathematical and astronomical contexts and over a period of 700 years at the very most.

The system uses only two basic numeral-signs, the vertical wedge  for 1 and the corner-wedge or *Winkelhaken*  for 10, to write any number between 1 and 59. Normally, this meant that small numeral-phrases were identical to those of the Sumerian cuneiform system. Nevertheless, certain graphic changes (shown in Table 7.14) were made to the numeral-phrases for 4, 7, 8, 9, and 40, so that, instead of grouping signs in at most two rows of up to five signs, three rows of no more than three signs were used. This shift eliminated any phrases that placed four or five signs side by side, and may have increased the system's legibility (Powell 1972a: 16).

Table 7.14: Graphic changes in numeral-phrases

	4	7	8	9	40
Sumerian					
Babylonian					

Unlike earlier Mesopotamian systems, all of which were primarily cumulative-additive, this system is cumulative-positional, combining the two basic signs in multiple positions to express exponents of 60. It is thus a base-60 system with a sub-base of 10. It has an additive structure within each exponent, because of the way that 10-signs and 1-signs combine together, but a positional structure among different exponents. Just as in the Sumerian and Assyro-Babylonian systems, subtractive notation was used frequently to write numbers such as 9 (10 lal 1) or 19 (20 lal 1) (Thureau-Dangin 1939: 106). According to the rules of the system, 4,252,914 would be written as 
 $=(19 \times 60^3) + (44 \times 60^2) + (20 \times 60) + 14$.

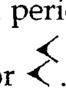
In addition to expressing integers, positional numerals could be used to express fractions using the sexagesimal fractional exponents of 60: 1/60, 1/3600, 1/216000, etc. Yet, during the Old Babylonian period, the positional numerals did not have any sign for zero to indicate an empty position within a numeral-phrase, nor was there any way to distinguish an integer from a fraction (i.e. there was no "sexagesimal point"). Powell (1976: 421) points out that many texts list numbers in columns in which the positional values of all the numbers are lined up with one another, in which case there is less possibility of misinterpretation. When numbers are embedded in the middle of a text or occur alone, the lack of a zero leads to ambiguity; there is no way, except through contextual information, to determine which positional value expressed which exponent, and thus a single numeral-phrase could have an infinite number of readings. The simple phrase 𒌋𒈫 𒐈 could mean 723 (12 x 60 + 3), 43380 (12 x 3600 + 3 x 60), 12.05 (12 + 3/60), and so on, depending on which positional values we assume are indicated.

When the empty position was medial (both preceded and followed by numerals), this difficulty was sometimes solved by using a large empty space to indicate the empty position (Neugebauer 1957: 20). Thus, 𒁹 𒎙 (80) could be distinguished from 𒁹 𒎙 (3620). Yet this technique was not used universally, and in some texts what looks to be a large space does not bear any numerical significance. Moreover, unless numeral-signs were arranged in columns, there was no way during the Old Babylonian period to distinguish numbers where the empty position came at the end or beginning of the numeral-phrase. Nevertheless, by organizing numbers in columns, and through common-sense interpretations of texts, Babylonian mathematicians would not have experienced insurmountable difficulties in reading numbers despite these ambiguities.

The date of origin of the positional system is still somewhat in doubt. Most texts containing the positional numerals are mathematical texts of the Old Babylonian tradition, and thus date between 2000 to 1600 BC, with the majority from the latter part of that period (Powell 1976: 419). There is some evidence that it may have occurred somewhat earlier, in the 21st century BC, during the Ur III (Neo-Sumerian) period. Powell (1972a: 14) notes that the late Sumerian system of weight units is purely sexagesimal and notated in a way that could be ancestral to positional notation. Powell (1976: 420) also found positional numerals on several early texts that led him to assert that the development of positional numerals occurred in the 21st century BC at the very latest. If this is the case, it may have been due to the considerable administrative reforms introduced in the Ur III period, which followed, in part, from managing much larger amounts of goods than had previously been the case (Høyrup 1985: 9; Powell 1976: 422). This hypothesis has not been accepted universally. Nissen, Damerow, and Englund (1993: 142) remain agnostic regarding Ur III positional numerals, because most texts can only be dated paleographically and the numerals do not show much variation throughout time. I am quite unconvinced by Whiting's (1984) assertion that the positional numerals developed as early as the Old Akkadian period (i.e. the 24th or 23rd centuries BC).

The Old Babylonian texts that contain positional numerals are all of a mathematical character. These range from simple multiplication tables and arithmetical exercises to complex problems that can legitimately be called algebra – problems that were useless for administration and approach what we might call pure mathematics. Many arithmetical exercises and texts for translating numerals into the new positional system date from the Old Babylonian period, indicating the existence of a vigorous process for teaching the system to scribes (Nissen, Damerow, and Englund 1993: 142-147). Yet, because non-mathematical texts did not contain positional numerals, scribes who did not write mathematical texts probably would not have been familiar with the positional numerals and would have written numbers only in the Assyro-Babylonian or Sumerian systems. The converse is not the case, however; Neugebauer (1957: 17) notes that mathematical texts containing many positional numbers are often dated in Assyro-Babylonian numerals, indicating that the positional systems' users also knew the common system. There is little evidence that the numerals were actually used to perform arithmetical calculations, as opposed to writing down results, although Powell (1976: 420-421) has found evidence of arithmetical clay tablets being moistened and re-used, from which he speculates that calculations were made on 'scratch pads' that could then be rewritten to record results after erasing the preliminary work.

After the end of the Old Babylonian period around 1600 BC, the positional system apparently ceased to be used for over a millennium. Because we have enormous textual evidence from the intervening period, it is extremely unlikely that the system was in continuous use during the interim. Rather, when the positional numerals re-emerge in the Seleucid period (the beginning of which is dated from the Alexandrine conquest of 332 BC), it must have been as a deliberate revival of the system of an older period (though we know little of the specific reasons for the system's reappearance). The numerals from this period, while largely similar to the Old Babylonian ones, differ in several interesting ways. First, subtractive expressions such as "20 *lal* 1" for 19 are no longer used in Seleucid texts (Neugebauer 1957: 5). Second, while in the Old Babylonian period the positional system was used exclusively in mathematical texts, by the Seleucid period it was used in astronomical texts as well (Neugebauer 1957: 14).

The most important change was the introduction, in certain circumstances, of a sign for zero to fill in an empty position within a numeral-phrase. Just as in the Old Babylonian period this system was the first positional system ever developed, so in the Seleucid period it also provides us with the first example of a zero-sign, usually written as  or . This sign could be used at the beginning of a numeral-phrase to indicate that the ones place was empty (i.e. to distinguish a fraction from an integer) or in a medial position to prevent misreading 3620 as 80, as above (Neugebauer 1957: 20). Much more rarely, zero could be used at the end of a numeral-phrase; Ifrah (1998: 153) cites a few instances, but in most cases, the empty final position had to be determined contextually.

Neugebauer (1941) emphasized that the primary role of the zero-sign was as much epigraphical as it was mathematical. He demonstrated that in a small number of texts, a "zero-sign" was inserted where none was warranted, apparently superfluously, but that in fact, all of these signs occurred in numeral-phrases preceded by an amount in tens and followed by an amount in units. This was done in order to preclude misreading 𒎙 𒐌 (20x60 + 7, or 1207) as 𒎙𒐌 (27); by writing the former as 𒎙𒑱𒐌, the latter interpretation is prohibited. In such numeral-phrases, the "zero-sign" does not indicate an empty position, but simply separates two consecutive positions (Neugebauer 1941: 213). In fact, the zero-sign was originally used to separate sentences. We should be cautious in drawing any historical implications regarding the development of zero out of a separator-sign, however. We must also be careful not to assume from the existence of a sign for zero that the Babylonians conceived of zero as an abstract number. I am aware of no Babylonian text that contains the bare numeral-phrase 𒑱; it always occurs in phrases with other signs. Thus, 𒑱 was not equivalent to 0 in the same way that 𒎙 was equivalent to 20. This more abstract concept of zero accompanied by a special sign for that concept developed independently among the Greeks and Indians, but probably never among the Babylonians.

The importance of the Babylonian positional numerals lies solely in their use in mathematics and astronomy. Despite the use of the positional principle, the system was restricted to an extremely small group of Babylonian scholars in both the Old Babylonian and Seleucid periods. It does not appear to have been known by merchants or administrators, and certainly did not diffuse to other peoples of the Middle East, such as the Aramaeans, Phoenicians, and Persians. There is no evidence to warrant the system's survival as late as 200 AD, at which time Menninger (1969: 398-9) believes the Mesopotamian system to have been borrowed into Brāhmī (!), thus leading to the development of Indian positional numerals (see also Février 1948: 585). Such arguments do not hold up to the facts, especially given the striking differences between the two systems, one of which is sexagesimal and cumulative-positional, the other decimal and ciphered-positional.

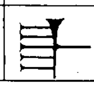
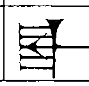
The only direct descendant of the Babylonian positional numerals is the sexagesimal Greek positional system used by classical mathematicians and astronomers to represent fractions. Around the 2nd century BC (by which time Babylonia was firmly under Seleucid control), the sexagesimal system was borrowed by the Greeks and combined with their alphabetic numerals to produce a ciphered-positional, base-60 numerical notation system (ch. 5). Neugebauer (1975: 590) states that the use of a sexagesimal division of the circle into 60 parts by Eratosthenes (ca. 250 BC) is the earliest evidence for this borrowing, although Eratosthenes did not use the sexagesimal fractions. I tend to agree with Ifrah (1998: 156) that it was not until the 2nd century BC that the Greek sexagesimal-positional fractions developed. It should be remembered, in any case, that the Greek system was used only in mathematical and astronomical texts, and *only* for fractions. The quasi-positional cuneiform system used in a few texts in the city-state of Mari also appears to derive partly from the Babylonian positional system.

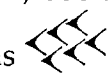
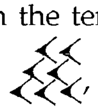
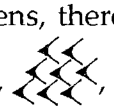
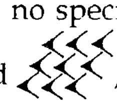
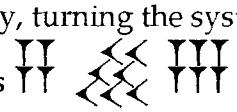
The Babylonian positional system survived somewhat longer than its Assyro-Babylonian counterpart did, partly because it was used in a limited set of contexts. Nevertheless, while the Seleucid astronomical texts are important from the perspective of the history of science, they clearly represented the work of a limited group of scholars whose knowledge was being surpassed by Greek mathematics and astronomy even in the 4th century BC. The Greek alphabetic numerals were those used for everyday purposes as well as for mathematics, and so it was quite natural that the Babylonian system, whatever its merits, would fall into decay. The last example of positional cuneiform numerals dates from the 1st century AD (Powell 1972: 6a). That this system, the first positional system ever and one much-lauded by modern scholars, should have fallen into decay so quickly (having already been abandoned once before by its own inventors) suggests that the advantages of positional systems do not correlate closely with their survival.

Mari

The city of Mari, located on the Euphrates River at the border of modern Syria and Iraq, was an independent city-state between the 20th and 18th centuries BC. During this time, it was engaged in extensive trade relations with Canaan and Babylonia. A large number of cuneiform tablets have been recovered from Mari, mainly dating to the 18th century BC, upon which a very unusual numerical notation system has been found. This system is shown in Table 7.15 (Durand 1987; Ifrah 1998: 142-146).

Table 7.15: Mari numerals

1	10	100	1000	10000
				

Below 100, the system is purely cumulative-additive. For the units, it is identical to the Assyro-Babylonian system, but in the tens, there is no special sign for 60; rather, the higher decades are written as , , , and , respectively. For the hundreds, the multiplicative ideogram  (ME) was often used (preceded by unit-signs for multiples). A unique feature of many mathematical texts from Mari is that they omit the  sign entirely, turning the system into a quasi-positional one. Thus, 476 is written in one inscription as  (Soubeyran 1984: 34). The sign for 1000 is a syllabic representation of "LI-IM", while the sign for 10,000 is created by superposing the signs for 10 and 1000. For exponents above 1000, the system is multiplicative-additive; there are no instances where higher exponent signs are omitted. The major differences between the Mari system and the Assyro-Babylonian system are thus the lack of a sign for 60 and the occasional use of positional notation for the hundreds exponent. The omission of the multiplicative sign for 100 is very interesting, but it does not represent a fully positional system. If it were, we would expect 476 to be written as (4 7 6), not (4 70 6). In this sense, the Mari system is similar to the experiments with positionality used in India in the 7th century AD, before the positional principle was fully understood (ch. 6).

The Mari texts, as a whole, do not contain only the system presented above. Administrative and commercial texts are usually written in the Assyro-Babylonian common system, while mathematical texts mostly use the Babylonian positional numerals. We thus know that the Mari scribes understood these systems perfectly well. What may have happened, at least in the case of mathematical texts, is that the aberrant system was used unofficially (perhaps for calculation) and then retranscribed for official purposes using the sexagesimal positional numerals (Soubeyran 1984: 34). It is quite possible that the decimal structure of the Mari numerals was borrowed from the Assyro-Babylonian ones, and that the idea of using positional notation for the hundreds was taken from the Babylonian positional numerals. This system should be considered as an aberrant and short-lived experiment with positionality, as the conquest of Mari by Hammurabi in 1755 BC ended its use.

<u>Hittite cuneiform</u>

In Chapter 2, I discussed the Hittite hieroglyphic system, which was probably borrowed from the Egyptian hieroglyphic system or the Linear B (Mycenaean) system. A separate Hittite script, written in cuneiform characters and related to the various Mesopotamian scripts, was used at the Hittite capital of Hattusha between the 17th and 13th centuries BC. The numerals used in this script are identical to those of the Assyro-Babylonian common system, except with certain graphic variations. They are shown in Table 7.16 (Rüster and Neu 1989: Table 7).

Table 7.16: Hittite cuneiform numerals

1	10	60	100	1000	10000
		 or 			

The system is decimal and cumulative-additive below 100, while the multiplicative principle is used for higher exponents. Numeral-phrases are written from left to right. To write 60 or multiples of 60, the Akkadian loanword *šu-ši* () was employed, just as it could be in other cuneiform systems. Yet, when writing numeral-phrases between 70 and 99, a simple vertical wedge represented 60 (Rüster and Neu 1989: 271). The sign for 100 is simply the ME syllable of Assyro-Babylonian numerals borrowed into Hittite. The complex and rare signs for 1000 and 10,000 appear to be unique to Hittite inscriptions. Thus, 7169 was written as  (7x1000 + 1x100 + 60 + 9) (Rüster and Neu 1989: Table 7). There is no evidence for the use of subtractive notation in the Hittite cuneiform numerical notation system.

Because the royal archives at Hattusha are our main source for Hittite cuneiform inscriptions, we do not yet know a great deal about the range of functions for which Hittite numerals were used. We can be quite certain that the numerals were borrowed from the Assyro-Babylonian additive numerals, which were widely used at the time the Hittite numerals are first attested, given the extreme closeness in structure and numeral-signs between the two systems. There does not appear to be any connection between the Hittite cuneiform and the Hittite hieroglyphic numerals, which are entirely different in structure. With the collapse of Hittite power in the 13th century BC, the cuneiform numerals ceased to be used.

Ugaritic

Very little is known about the numerals that accompanied the Ugaritic script, which was used between the 15th and 12th centuries BC at Ugarit on the Mediterranean coast. Gordon (1965: 42) reports that the cuneiform ideograms for 1 (𒁹) and 10 (𒌋) were used in various administrative documents, but notes that Ugaritic numerals were normally written lexically. We have no idea whether higher numbers could be expressed through numerical notation. Numeral-phrases, like the script itself, were written from left to right. Février (1948: 577) indicates that they were "empruntés au monde suméro-akkadien, mais légèrement modifiés", but does not provide any evidence for this assertion. Presumably, by the 15th century BC, the source system from which the Ugaritic numerals were borrowed would have been the Assyro-Babylonian common system rather than the Sumerian system, which was obsolete by that time.

Old Persian

The script that is now known as Old Persian was invented early in the domination of the Achaemenid Empire over Mesopotamia, probably near the beginning of the reign of Darius I (522 to 486 BC) (Testen 1996). It is an alphasyllabary; thus, while its letter-signs are cuneiform, it represents a distinct break from the older Assyrian and Babylonian scripts. Nevertheless, the Old Persian numerals clearly indicate their descent from the Assyro-Babylonian common numerals in both signs and structure. The numerals are shown in Table 7.17 (Testen 1996: 136).

Table 7.17: Old Persian cuneiform numerals

1	10	100
𐏑	𐏓	𐏕

The system is decimal and cumulative-additive, and numeral-phrases are written from left to right. We have no inscriptions that show how the higher hundreds values were formed; it is possible that the horizontal bar above the vertical wedge for 100 is a multiplicative form, but this cannot be confirmed. Whereas the Assyro-Babylonian units could be grouped in sets of two or three (e.g. 8 = 𒐍, Old Persian units and tens were arranged in at most two rows, with odd units represented at twice the size of paired ones. Thus, 79 might be written as .

Despite the lack of multiplicative higher exponents, a special sign for 60, and other features of the Assyro-Babylonian system, the invention of the Old Persian numerical notation system was a result of contact with Semitic-speaking Assyrians and Babylonians in the years following the Persian conquest. There are a number of Babylonian/Persian bilingual inscriptions, and we know that the two systems existed side by side at that time. Yet, despite being used on Old Persian inscriptions in the late 6th and 5th centuries BC, these numerals were never especially popular. By the time that they were developed, the Aramaic numerals were the system used for international communication and commerce throughout Mesopotamia. The Old Persian system certainly did not survive the Alexandrine conquest of Persia.

Summary

The commonalities among the Mesopotamian numerical notation systems are limited to their use of cumulative notation with signs for 1 and 10 (with the exception of some of the metrological systems). In other respects, there is considerable variation among these systems, whether due to linguistic (Sumerian vs. Semitic) or functional (administrative vs. mathematical) factors. The survival of the sexagesimal base over a period of nearly 3500 years is testament to the Babylonians' archaic preservation of Sumerian traditions, but most of the numerals used after 2500 BC are primarily decimal. The mathematical functions of the various systems, while interesting to historians of mathematics, are minimal in comparison to their administrative and literary functions.

The conquest of Mesopotamia by the Achaemenids and later the Seleucids sounded the death-knell for native Mesopotamian traditions such as the numerals, and Aramaic and then Greek numerals came to be used for most purposes. Even the positional numerals, the hallmark of Babylonian arithmetical achievement, quickly disappeared under conditions of cultural and political domination, demonstrating the relevance of social factors to the survival of numerals. The Assyro-Babylonian common system, in particular, was borrowed and modified in regions of the Middle East where Mesopotamian influence was strong. Yet the history of Mesopotamian numerals is linear rather than branching, with each system giving rise to its successor but not giving rise to many systems outside Mesopotamia. Multiple systems were often employed at the same time within Mesopotamia, the use of which was divided by their context in ways that remain unclear. With the sole exception of the Levantine family, which is derived from both this family and the Hieroglyphic family, the Mesopotamian family of systems did not give rise to a large number of descendants either within Mesopotamia or without.

Chapter 8: East Asian Systems

The East Asian family of numerals is the final Old World family of this study. The East Asian numerical notation systems, like the region's scripts, reflect the pervasive importance of Chinese civilization over the past three millennia. The 'classical' Chinese numerals used from the Qin dynasty (221 - 206 BC) to the present day, and which spread into Japan, Korea, and Vietnam, as well as China, are foremost in duration and significance in this family. Still, the history of East Asian numeration is one neither of total Chinese hegemony nor of complete stasis. In fact, the systems of this family are much more diverse in structure than are those of any of the other families I have investigated. I identify these systems as part of the same family from known historical connections, as well as similarities in their numeral-signs. Table 8.1 indicates the most common numeral-signs of the East Asian systems.

Table 8.1: East Asian numerical notation systems[1]

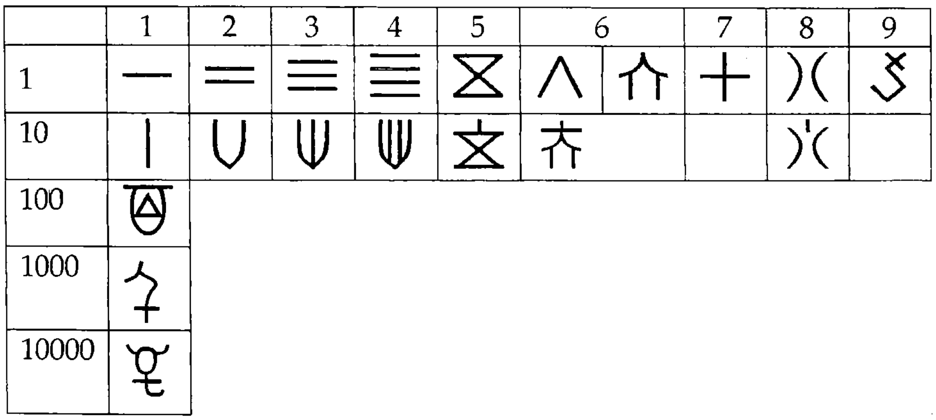
	1	2	3	4	5	6	7	8	9	10	100	1000	10000	0
Shang / Zhou	一	二	三	亖	[illegible]	[illegible]	十	)(	[illegible]	|	[illegible]	[illegible]	[illegible]	
Chinese classical	一	二	三	四	五	六	七	八	九	十	百	千	萬	零
Rod-numerals	𝍠	𝍡	𝍢	𝍣	𝍤	𝍥	𝍦	𝍧	𝍨	𝍩	𝍠	𝍩	𝍠	
Late rod-numerals	𝍠	𝍡	𝍢	╳	[illegible]	𝍥	𝍦	𝍧	[illegible]	𝍩	𝍠	𝍩	𝍠	〇
Chinese commercial	〡	〢	〣	〤	〥	〦	〧	〨	〩	〸	[illegible]	[illegible]	[illegible]	〇
Chinese positional	一	二	三	四	五	六	七	八	九	十				〇
Kitan	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]			
Jurchin	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	
Ryukyu	•	••	•••	••••	ㄴ					十	○	⊙		

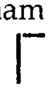
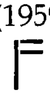
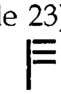
[1] Not all numeral-signs shown; see individual entries for complete inventories of signs.

Shang and Zhou

The first attested writing in East Asia dates to the latter part of what we now call the Shang Dynasty (ca. 1523 - 1028 BC). The most common Shang inscriptions are records of royal divinations, called "oracle bone inscriptions" by modern scholars. These brief texts, which date from 1300 to 1050 BC, often contain numerical indications of tribute received, animals hunted, or numbers of sacrificial victims (Takashima 1985: 45). The attested numeral-signs used on oracle-bone inscriptions are shown in Table 8.2 (Needham 1959: Table 22, 23; Djamouri 1994: 39).

Table 8.2: Shang numerals



The Shang numerical notation system combines the nine unit-signs with signs for the exponents of 10; it is thus multiplicative-additive and decimal. The numeral-signs for 1-4 are cumulative combinations of horizontal strokes, while the signs for 10-40, only slightly less obviously, are ligatured combinations of vertical strokes; 20, 30, and 40 are never expressed using multiplicative expressions involving the signs for 2, 3, and 4. Another apparent irregularity is found on certain inscriptions indicating months of the year; Needham (1959: Table 23) lists the signs for 11, 12, and 13 in these cases as being written as , , and . These can easily be analysed as perfectly regular combinations of the sign for 10 and various unit-strokes.

For the tens between 50 and 90, the unit-sign is placed below the sign for 10, which was normally a vertical stroke but could apparently be a cross when writing 60. For the hundreds, the unit-sign was placed above the exponent-sign, while for the thousands and ten thousands, the relevant unit-sign was superimposed upon the exponent-sign. These latter two exponent-signs are identical to the Shang characters for 'man' and 'scorpion', respectively (Djamouri 1994: 15-16). Numeral-phrases were written in vertical columns read from top to bottom, with the highest exponent at the top. Figure 8.1 shows how the number 4539 might be written in a Shang inscription.

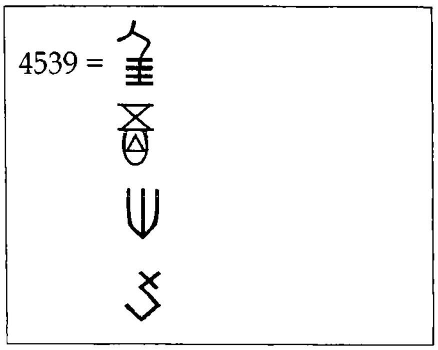


Figure 8.1: 4539 in Shang numerals

In almost all the oracle-bone inscriptions, however, numeral-phrases are not found alone, but are accompanied by a character for the object being quantified. On this basis, Djamouri regards Shang numeral-phrases as determinatives of noun-phrases, and argues that each sign was read as a single morpheme in the ancient Chinese language (Djamouri 1994: 33). It is significant that each numeral-sign corresponds with a single Chinese morpheme; this correspondence between language and numerals is atypical, and leads Djamouri to regard the Shang numerical notation system as a purely linguistic rather than a 'graphic' phenomenon. This feature, which it shares with the Chinese classical system, raises the issue of whether we ought to consider such quasi-lexical formulations to be "real" numerical notation systems. I will return to this issue at the end of this chapter.

Needham insists that the Shang numerals contain "place-value components", because the unit-signs are used in combination with the higher exponent-signs, and on this basis contends that the system was "more advanced and scientific than the contemporary scripts of Old Babylonia and Egypt" (Needham 1959: 13). This is an artifice created by his insistence that the exponent-signs are not numerals, but are instead non-numerical indicators of place-value (1959: 14). Because the exponent-signs on their own (without a stroke to indicate one multiple of the exponent) do not constitute numeral-phrases, Needham insists that they are somehow not numeral-signs. The concept of a non-numerical sign indicating place value is paradoxical, since place value by definition indicates numerical value by position and not by graphic signs. Unlike ancient positional systems, such as the Babylonian sexagesimal numerals, the value of a Shang unit-sign was always combined with an exponent-sign to represent higher numbers. It had no sign for zero; if a particular exponent was not needed, no sign indicated its absence in the numeral-phrase.

The origin of the Shang numerical notation system is quite clearly understood; like the Shang script, it was an independent invention. Needham (1959: 149), who frequently favours diffusionistic arguments, suggests that the only potential ancestor for the Shang system is the Babylonian astronomical (positional) numerals, and even then it could only have arisen through stimulus diffusion. Yet he is very skeptical of the likelihood of any connection between the two systems. Since the only solid evidence for such a transmission is his belief that the Shang and Babylonian systems share a common 'place-value component', I am entirely unconvinced by this theory.

Moreover, the correspondence of numeral-sign and number-word suggests that the Shang numerals have a linguistic origin. If the signs originated, as suggested by Djamouri, to represent perfectly morphemes in the language of their speakers, this further confirms their indigenous development. There are few resemblances between Shang numeral-signs and phonetic characters that suggest that those signs were chosen because the lexical numerals were homonyms of other words (Djamouri 1994: 18-19). Other theories of the origin of the numeral-signs include a) that they are derived from Neolithic pottery-marks, such as those found at Banpo (4800 – 4200 BC); b) that they are related to an ancient (and hypothetical) system of finger-numbering; and c) that they are related to the rod-numerals discussed below (Djamouri 1994: 18-22). None of the latter hypotheses needs to be given much attention, as they all require special pleading to be remotely plausible. I think it likely that the signs are of a mixed abstract and phonetic origin; more important than phonetic correspondences may be the fact that most of them are graphically quite simple as compared to the other Shang characters.

After the collapse of the Shang dynasty, large parts of what is now China were controlled by the Zhou dynasty, first from its western capital at Hao (1027 - 770 BC) and, after the failure of the Western Zhou state, by a more decentralized polity centered farther to the east at Luoyang (770 - 256 BC). The Zhou kingdoms continued to employ the script and numerals of the Shang. In the early Zhou period, oracle-bone inscriptions continued to be written, but from the 10th to the 3rd centuries BC, Zhou numerals often were stamped on bronze vessels and coins, and were inscribed in texts of various sorts (Needham 1959: 5). The increasing complexity of Chinese society over this long period brought the numerals into use for a much wider range of functions than is documented to have previously been the case. While the Zhou numerals are structurally identical to the earlier Shang numerals, the numeral-signs are slightly different, as shown in Table 8.3 (Needham 1959: Table 22, 23; Djamouri 1994: 39).

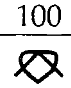
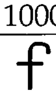
Table 8.3: Zhou numerals

	1	2	3	4	5	6	7	8	9
1	一	二	三	亖	𠄡	介	十	)(	九
10	丨	廿	卅	卌	𠄡十	介十	Unattested	)(十	九十
100	百								
1000	千								
10000	萬								

After the collapse of the Western Zhou state in 770 BC, China became politically fragmented. While the Eastern Zhou continued to rule a smaller area, the Spring and Autumn (770 - 480 BC) and Warring States (480 - 221 BC) periods featured great social disorder, but also great creativity in literature and philosophy (most notably in the development of Confucianism). Chinese numeration increased exponentially in the variability of numeral-signs used during this period. No comprehensive study of these numerals has been undertaken; however, Pihan (1860: 10) provides a comprehensive chart showing the various numeral-signs used between the 6th and 2nd centuries BC, in which no fewer than 38 different signs for 10,000 are listed. Despite the extraordinary paleographic variability during this period, the signs shown in Table 8.3 continued to be the ones most commonly found on coins and bronzes until the 3rd century BC.

Among the variant signs developed in the late Zhou and Warring States periods are the set of exponent-signs that are the immediate ancestors of the corresponding Chinese classical numerals. Structurally, this system is identical to the earlier one, except that the sign for 10 could also combine multiplicatively with the unit-signs for 2 through 4. These signs, shown in Table 8.4, were used between the 6th and 3rd centuries BC (Needham 1959: Table 23). By comparing them with the exponent-signs in Table 8.3, it can be seen that there is little similarity between the two sets.

Table 8.4: Late Zhou exponent-signs

10	100	1000	10,000
十			万

Just as there is no sharp break in the forms of signs between the Shang and Zhou systems, neither is there a distinct break between the Eastern Zhou / Warring States numerals and those of the Qin dynasty; rather, the former gradually transformed into the latter. Yet, given the rather important changes in Chinese writing that took place after the unification of the country in 221 BC, I have chosen that point of demarcation to separate the earlier numerical notation system from the 'classical' Chinese system.

Chinese rod-numerals

Before turning to the classical Chinese system, however, I will address a system that developed alongside the written numerals of the Warring States period. This system, known in Mandarin as *suan zi* and in English as 'rod-numerals', is peculiar in that it is not only a numerical notation system, but also a computational technology. In fact, though it is poorly known in the West, rod-numeral calculation was the primary computational technology used in East Asia before the 16th century, when the bead-abacus supplanted it. The standard rod-numerals are indicated in Table 8.5 (Needham 1959: Table 23).

Table 8.5: Early rod-numerals

	1	2	3	4	5	6	7	8	9
1s	𝍠	𝍡	𝍢	𝍣	𝍤	𝍥	𝍦	𝍧	𝍨
10s	𝍩	𝍪	𝍫	𝍬	𝍭	𝍮	𝍯	𝍰	𝍱
100s	𝍠	𝍡	𝍢	𝍣	𝍤	𝍥	𝍦	𝍧	𝍨
1000s	𝍩	𝍪	𝍫	𝍬	𝍭	𝍮	𝍯	𝍰	𝍱

The system is quite simple to learn and use; vertical and horizontal lines are sufficient to write any number. Normally, for the ones position, vertical strokes signify 1 and horizontal strokes signify 5; combinations of vertical and horizontal strokes indicate the value of the units position. Conversely, for the tens, the values of the individual strokes are reversed, so that horizontal strokes mean 1 and vertical strokes mean 5. Each successive position is modeled alternately on the ones and the tens; positions in which the sign for 1 is vertical (ones, hundreds) are called *zong*, while those in which it is horizontal are called *heng* (Needham 1959: 8-9). In the earliest rod-numerals (4th century BC to 3rd century AD), the use of *zong* and *heng* numerals as appropriate to their position was not strict, so that horizontal strokes could be used for ones and vertical strokes for tens. However, the system stabilized by the end of the Han dynasty. No zero-sign was used at this early date; rather, the numeral-signs were lined up strictly by position, leaving blank spaces as appropriate, obviating the need for a zero.

The rod-numerals constitute a cumulative-positional system with a base of 10 and a sub-base of 5. While it is possible to regard each sign - such as 𝍨 for 9 - as a single sign, thus making this system ciphered-positional, the system's true structure is best reflected by classifying this system as intraexponentially cumulative, which allows us to recognize how the sign is constituted and to note its sub-base. Note that while the numerals 6 through 9 are written using compounds 5 and 1 through 4, the sign for 5 alone is always five strokes; if a horizontal stroke were used for 5 in a *zong* position, there might be more risk of confusion with the horizontal stroke for 1 in the next highest (*heng*) position. By virtue of the fact that the direction of the strokes alternates with each successive position, the rod-numerals are irregularly positional, since a given sign does not take its meaning solely from its position, but also from the orientation of the symbols within each position. To put the sign 𝍰 in the tens position indicates 70, but to put it in the ones or hundreds position would have violated the system's structure, except during the earliest phase of its history.

The rod-numeral system is infinitely extendable by using these two alternating sets of numeral-signs in successively higher positions. Decimal fractions could be written by designating one of the places as the "units" position, with the places to the right of that one representing 0.1, 0.01, etc. (Volkov 1994: 81). In numeral-phrases containing both whole and fractional positions, the ones position could be identified by the presence of a character beneath it to indicate what sort of thing is being counted (Libbrecht 1973: 73). Where numbers were arranged strictly by columns, however, it was not necessary to include this extra sign. Moreover, the system is even more flexible. In addition, as early as the Han dynasty, negative numbers could be written, either by using different-coloured rods (red for positive numbers, black for negative numbers) or by placing an extra rod diagonally across the last non-zero digit of the numeral (Lam 1986: 188). Figure 8.2 shows several numeral-phrases written in rod-numerals.

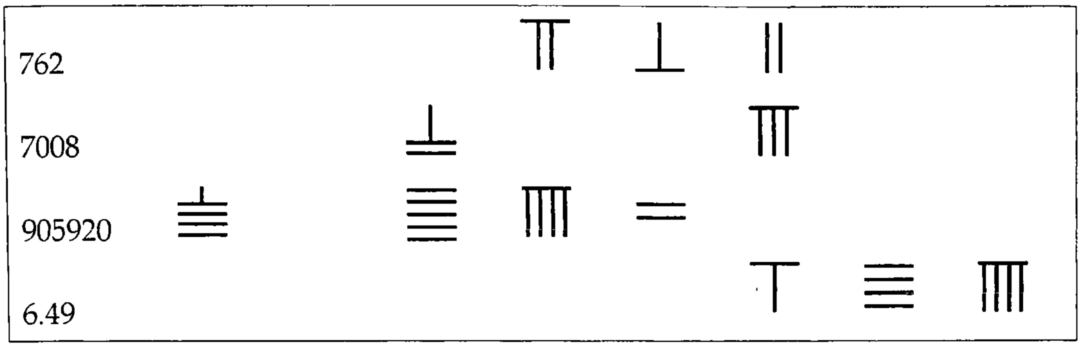


Figure 8.2: Rod-numerals (examples)

The rod-numerals have their origin, not in writing, but as a form of computation on a flat surface. Rods known as *chou* or *suan* were used to produce the signs shown in Table 5. While most rods were made of bamboo, others were made of bone, wood, paper, horn, iron, ivory, or jade (Lam 1987: 369). The earliest physical rods to be unearthed are several found at Fenghuangshan in Hubei province, which date to the reign of Wen Di (179 - 157 BC) (Mei 1983: 59). Textual and epigraphic evidence shows, however, that the rod-numerals were developed much earlier. Coins from the Warring States period frequently contain rod-numerals, so the system can hardly have been developed much later than 400 BC (Needham 1959: 5). Yet its acceptance was not automatic. The *Daodejing* (Tao Te Ching), written in the early 3rd century BC, advises that, "Good mathematicians do not use counting-rods", confirming that the system was in use at that time, while also showing that the system had not yet acquired the acceptance that it later would (Needham 1959: 70-71).

While the rod-numerals' origin as a means of computation is clear, the late Zhou numerals also may have influenced on their development. While the systems are of a very different structure, their signs are similar; the Zhou sign for 1 is a horizontal line and the sign for 10 a vertical line with a dot. Because the early rod-numerals did not have a regular orientation, a horizontal rod could indicate 1 and a vertical rod 10. Given that the inventor(s) of the rod-numerals were probably literate, they would have been familiar with the Zhou signs and may have borrowed them, though this explanation for the similarity in the signs cannot be proven.[2] The rod-numerals' cumulative-positional structure and quinary sub-base are both useful features that allow a limited number of rods to express any number. In practice, the use of physical rods would have limited the number of positions that could be managed easily, but this difficulty did not exist in the written rod-numerals. Thus, although the rod-numerals are identical in structure to the Greco-Roman abacus (which predates the rod-numerals by at least two centuries), I attribute this similarity solely to the common function served by the two technologies.

Lam Lay-Yong (1986, 1987, 1988) hypothesizes that the rod-numerals were ancestral to the Hindu positional numerals. Her evidence for this hypothesis is that the rod-numerals are positional and decimal, and there was considerable cultural contact between China and India in the 6th century AD, around the time when positionality developed in India. Because the rod-numerals were used in computation and commerce, she asserts that it is inconceivable that the Indians would not have learned of this system from the Chinese, and, since it is so practical, they obviously would have borrowed it (Lam 1988: 104). From this, she asserts that the rod-numerals are the ultimate ancestor of the Western numerals.

While Lam's hypothesis is plausible, I am deeply skeptical of its validity. Two immediate objections are that the Indian positional numeral-signs are those of the earlier Brāhmī numerals, not of the rod-numerals, and that the rod-numerals have no zero-sign (whereas the Indian system does). To the first objection, Lam responds that "since six of the nine digits in rod numeral notation were strange to them, they would naturally have

---

[2] I am highly dubious of the 4th and 3rd century BC coins that de Lacouperie (1883: 311-314) claimed to have found, inscribed with a mixture of Zhou numerals and rod-numerals using a ciphered-positional system and a circle for zero. Hopkins (1916) was stymied by this evidence, and Needham (1959) does not mention it. However, if it is genuine, and if de Lacouperie has interpreted it correctly, this would be a crucial piece of evidence helping to prove this hypothesis.

preferred their own numerals" (Lam 1986: 193). The notion that the rod-numerals were so foreign to the Indian mind as to require the total abandonment of its signs is unacceptable; who cannot comprehend the use of vertical and horizontal strokes? To the question of the zero, Lam replies that the abandonment of the alternating *zong* and *heng* positions required that the Indians develop a sign to fill the blank space (Lam 1986: 194). I do not think this follows; a blank space would have served just as well as a zero-sign in either system, and if the abandonment of the alternating positions created such difficulty, why would the Indian mathematicians have done it? Even more damaging to Lam's argument are two structural differences between the rod-numerals and the Indian numerals that she ignores entirely: the rod-numerals have a quinary sub-base that the Indian numerals lack, and the rod-numerals are intraexponentially *cumulative* whereas the Indian positional numerals are ciphered. Moreover, no Indian texts of the period mention rod-numerals or any other Chinese numeration. Indeed, as I will discuss below, the Indian positional numerals were seen as remarkable in China in the early 8th century AD, suggesting that the Chinese traders who hypothetically transmitted the rod-numerals to India were entirely unaware of the result of their transmission. Lam's theory is so weak that it is equally plausible that the Greco-Roman counting board, which was also quinary-decimal, cumulative-positional, and used in the Middle East, was an ancestor of the Indian numerals - that is, it is not very plausible at all.

In the 6th century AD, the numerals and the related rod-computation technique were introduced into Japan at a time when Chinese cultural, religious, and political influence in Japan was enormous; they were known in Japanese as *sangi* (Menninger 1969: 368). There is no evidence of their use outside China, Japan, and Korea. Around the same time, in China, the numerals, which had been used for nearly a millennium on coins, were replaced by the classical Chinese numerals. The last coins to use rod-numerals are the 5 *chu* coins of the Liang dynasty (502 - 557 AD), but these numerals are highly irregular (de Lacouperie 1883: 316-317). They continued to be written in Chinese texts and used directly for computation.

In the 12th and 13th centuries (the late Song dynasty), the rod-numerals transformed significantly from their original form. Although this was a time of considerable political turmoil in China, due to invasions by groups such as the Jurchin and Mongols, it was also a time of considerable scientific achievement. Table 8.6 indicates the system as it was used at that time (Needham 1959: Table 22; Libbrecht 1973: 68).

Table 8.6: Late rod-numerals

	1	2	3	4	5	6	7	8	9	0
zong (1, 100, ...)	𝍠	𝍡	𝍢	𝍣 ╳	𝍤 Ō	𝍥	𝍦	𝍧	𝍨 ⊼	○
heng (10, 1000, ...)	𝍩	𝍪	𝍫	𝍬 ╳	𝍭 ȯ	𝍮	𝍯	𝍰	𝍱 ⚹	○

The rod-numerals underwent three major changes around this time, all of which applied only to the writing of rod-numerals in texts, never to the physical manipulation of the rods, which apparently did not change. First, while the original (cumulative) signs for 4, 5, and 9 were retained, additional signs were introduced for those numbers. Because the only signs to change were those in which four or five cumulative strokes had previously been required, it is probable that the change was in part undertaken to simplify the signs, even though it meant that the system as used in texts differed from that used with physical rods. Structurally, these changes to the numeral-signs made the system less cumulative than it previously had been, though it was obviously still positional. Second, during this period, written numeral-phrases sometimes were condensed into single glyphs, compressing the individual signs together so that they formed a monogram. Needham (1959: 9) attributes this development to the requirements of the new technology of printing books. Thirdly, and perhaps most significantly, a circle was introduced as a sign for zero. The first text known to use a zero-sign is the *Shu shu* *jiu zhang* (Mathematical treatise in nine sections) of Qin Jiushao, published in 1247 (Libbrecht 1973: 69).[3] While Needham suggests that the idea of a circle for zero may have been an endogenous development, based on the philosophical diagrams of 12th century Neo-Confucian scholars, I concur with the vast majority of scholars in concluding that this development was due to influence from India (Needham 1959: 10). We may never know, however, whether the exact route of transmission was through Southeast Asia, Tibet, or India proper. Figure 8.3 shows the cumulative effects of these three changes.

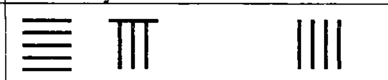
	Old style	New style
5804		

Figure 8.3: Early and late written rod-numerals

The characteristic of the rod-numerals that differentiates them from nearly every other numerical notation system is that their use was linked directly with arithmetical computation from the time of their invention. While they began as a system involving the physical manipulation of rods, they were rapidly adopted as a written numerical notation system by Chinese mathematicians. The earliest surviving mathematical text that discusses them is the *Jiu zhang suanshu* (Nine chapters of the mathematical art), written sometime between the 3rd century BC and the 1st century AD, probably in the latter half of this period (Lam 1987: 367-368; Volkov 1994: 81). After this point, most Chinese mathematical and astronomical texts until the 16th century used or discussed rod-numerals (sometimes accompanied by classical numerals).[4] In fact, most Chinese characters having to do with computation use the 'bamboo' radical due to its association with bamboo computing rods (Needham 1959: 72).

[3] As I will discuss below, this text is also the first to use a circular sign for zero in conjunction with the classical numerals.

[4] We may never know the true extent of their use, since many printers considered the rod-numerals, with their vertical lines, to be insufficiently literary, and replaced them with the classical numerals (Needham 1959: 8).

The introduction of the bead-abacus (*suan pan*) in the 14th or 15th century brought this novelty into direct competition with the rod-numerals. It is clear from textual sources that the *suan pan* was far more efficient for computational purposes than the rod-numerals, making their demise a foregone conclusion. The divorcing of rod-numerals from the physical manipulation of rods made their use in written form rather archaic. Throughout the Ming dynasty (1368-1644), they were used increasingly rarely in Chinese books, and were probably a historical curiosity by 1600 (Cheng 1925: 493). None of the many 17th and 18th century European scholars who mentions the abacus also notes the rod-numerals (Needham 1959: 80). However, rod-numerals continued to be used in Japan for some time after they had been abandoned in China, and were apparently not yet obsolete in the 18th century, when they were still used in some books (Menninger 1969: 368-9). A new Chinese computing technique developed in the 17th century in which computing rods were inscribed with numerals, probably under the influence of the system of numbered rods developed by the English mathematician John Napier in 1617 (Needham 1959: 72). This technique (similar to a slide rule) need be given no attention here, since it is not a numerical notation system but simply a computing technology that uses the Chinese classical numerals, and at any rate is very different from the rod-numerals of antiquity.

The manipulation of rod-numerals on boards appears to have been nearly as important to ancient and medieval Chinese scientific and commercial calculation as the bead-abacus would later be. The link between rod-numerals and computation is very unusual for numerical notation systems. Their origin and persistence must have had a great deal to do with their efficiency for this function. However, this supports rather than refutes my thesis that the history of numerical notation systems should be divorced from their use as mathematical tools. The rod-numerals and the classical Chinese numerals coexisted for nearly 2000 years, and yet the former had no noticeable impact on the latter. If there truly existed a unilinear trend for positional systems to supplant additive ones, we would expect either that the rod-numerals would replace the multiplicative-additive classical numerals entirely, or at least facilitate their transformation into a ciphered-positional system. Yet the use of positionality in conjunction with the Chinese classical numerals does not antedate the 13th century, and never displaces the older system entirely.

Chinese classical

The basic numerals associated with the Chinese script are perhaps the most stable symbol system presently in use; the numeral-signs of the Qin dynasty (221 - 206 BC) are practically identical to those used in modern Chinese literature. While there are structural differences between that system and the way the numerals are normally used today, ancient numeral-phrases are still easy to read. The basic numeral-signs used in this system are shown in Table 8.7.

Table 8.7: Classical Chinese numerals

1	2	3	4	5	6	7	8	9
一	二	三	四	五	六	七	八	九
yi	*er*	*san*	*si*	*wu*	*liu*	*qi*	*ba*	*jiu*

10	100	1000	10000	100,000,000
十	百	千	萬 or 万	亿 or 億 or 萬萬
shi	*bai*	*qian*	*wan*	*yi* or *wan wan*

In traditional writing, numerals, like the script, were arranged in columns from top to bottom, with the highest exponents first. In modern writing, normally numerals are written in rows from left to right, although right-to-left writing is not unknown, and in these cases right-to-left numeration is employed. The basic system is multiplicative-additive; numbers are written by combining the signs for 1-9 with the appropriate signs for the exponents of 10 to indicate their multiplication, and then taking the sum of these pairs of signs. There is no exponent-sign for the units; the unit-signs for 1-9 stand alone. When the value of a given exponent is zero, both the unit-sign and the exponent-sign are omitted. There is no zero-sign in the traditional system, although there is in modern Chinese numerals (which I will describe later in this section). In addition to these standard signs, there are three non-standard signs used for 20, 30, and 40 (廿 (*nian*), 卅 (*sa*), and 卌 (*shu*), respectively), which have their origins in the Shang/Zhou cumulative signs for the lower decades (Needham 1959: 13). These signs were used in literary contexts, particularly in poetry, for paginating certain texts, and when denoting days of the month. They are still used occasionally, although the sign for 40 is very rare because it is not needed to enumerate days of the month. It was always acceptable (and now is preferred in most contexts) to use the standard multiplicative combinations of the unit-signs 2 through 4 and the exponent-sign for 10.

No unit-sign is needed when the multiple of an exponent is 1, but the unit-sign for 1 is sometimes included in such cases. When writing the numbers 11 through 19, the unit-sign attached to 10 is always omitted, although in numbers such as 214 the unit-sign for the tens is sometimes included. In addition, when writing the basic exponents of the base (10, 100, 1000, etc.) the unit-sign is normally omitted, so that 1000 would be written without an exponent-sign but 1002 might have one.

Unlike Western numerals, which are grouped in chunks of three digits, Chinese numerals are grouped in sets of four, using the character *wan* (10,000, or, if you will, 1,0000) as a sort of meta-exponent (Mickel 1981: 83). Any number from 10,000 to 100,000,000 could be written by placing a multiplicative numeral-phrase from 1 to 9999 before the sign 萬 (10,000). The system did not stop there, however; multiples of 100,000,000 could be written by placing a multiplicative numeral-phrase in front of the signs 萬萬 (10,000 x 10,000) or by using a unique sign for 100,000,000, either 亿 or 億. Figure 8.4 indicates several numeral-phrases that reflect these rules in operation.

15	十五	10 5
118	百十八	100 10 8
74,002	七萬四千二	7 10000 4 1000 2
1,072,419	百七萬二千四百十九	100 7 10000 2 1000 4 100 10 9
4,703,600,854	四十七億三百六十萬八百五十四	4 10 7 100 mil. 3 100 6 10 10000 8 100 5 10 4

Figure 8.4: Chinese classical numeral-phrases

Another technique for expressing large exponents of 10, which developed very early in the history of Chinese numeration, involved a complex system of exponent-signs that was assigned three different series of values, as shown in Table 8.8 (Needham 1959: 87). These exponent-signs are combined with the nine basic unit signs multiplicatively, and thus are simply an extension of the basic system. These signs first appeared in the *Shu shu ji yi*, a text that dates to about 190 AD. While this system may seem hopelessly complex and ambiguous, this confusion is identical to that resulting from the different values assigned to *billion* and *trillion* in American and European usage. In the lower series, each exponent is one greater than the one before it; in the middle series, each exponent is four greater than the one before it, and in the upper series, each exponent is double the one before it. The first sign in all three series is the standard sign for 10,000, and the second sign (yi) is one of the basic signs for 100 million (thus corresponding with the middle and upper series, but not the lower one). [5] This system is still used occasionally.

[5] There is apparently much confusion among scholars regarding these signs and their values, as can be seen by comparing the lists of Pihan (1860: 3), Perny (1873: 100), Needham (1959: 87), and

Table 8.8: Chinese higher exponent-signs

Sign	Phonetic value	Lower series *xia deng*	Middle series *zhong deng*	Upper series *shang deng*
萬	*wan*	10^4	10^4	10^4
億	*yi*	10^5	10^8	10^8
兆	*zhao*	10^6	10^{12}	10^{16}
京	*jing*	10^7	10^{16}	10^{32}
垓	*gai*	10^8	10^{20}	------
秭	*zi*	10^9	10^{24}	------
壤	*rang*	------	10^{28}	------
溝	*gou*	------	10^{32}	------
澗	*jian*	------	10^{36}	------
正	*zheng*	------	10^{40}	------
載	*zai*	------	10^{44}	------

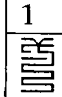
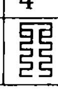
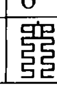
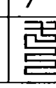
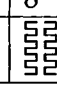
The Chinese numerals began to take their modern form starting in the 3rd century BC, developing directly out of the numerals used in the Warring States period, a process that would take several centuries to complete. With the spread of a unified administrative apparatus under the Qin and Han dynasties, the Chinese numerals spread throughout the region under direct and indirect imperial control. The unification of China led to many efforts to standardize the forms of Chinese script and numeral-signs, although this was not accomplished to any significant extent until late in the Han dynasty. At the same time as the signs of the system were being stabilized, however, Chinese writers began to use calligraphic variants and other modifications of the basic system for specific functions. These variants used different numeral-signs (ranging from mild paleographic variations to radically different signs), but their structure is identical to that of the basic system (decimal and multiplicative-additive).

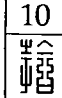
Ifrah (1998: 278). I have followed Needham's list, which seems to have been accepted most widely.

Perhaps the most important of these are the 'accountant's numerals' (*da xie shu mu zi*), which developed as early as the 1^{st} century BC (Needham 1959: 5, Table 22). Structurally, they are identical to the classical numerals, but while the classical numeral-signs are quite simple, the accountant's numerals were intentionally made very complex; thus they were considered more elegant and less susceptible to falsification. The signs are homophones of the phonetic values of the appropriate Chinese words, so they bear no graphic resemblance to the basic signs. Hopkins' (1916) analysis of their origin as phonetic variations of the standard numerals is dated but quite thorough. Despite their name, they were not used only for accounting but, for instance, were also used on 13^{th} century coins (de Lacouperie 1883: 318-319). Today, they are still used occasionally on cheques, banknotes, coins, and contracts in order to prevent falsification.

Another highly complex variant of the classical numerals are the *shang fang da zhuan*, a variant set of numeral-signs that developed in the Han Dynasty (Pihan 1860: 13; Perny 1873: 113). These numerals, which are structurally identical to the classical system, are highly stylized linear versions of the standard numeral-signs that were designed to be used on seals, and are still sometimes used for that purpose today. These signs are shown in Table 8.9.

Table 8.9: Shang fang da zhuan numerals

1	2	3	4	5	6	7	8	9
								

10	100	1000	10000
			

The transmission of the Chinese classical numerals has always been (and continues to be) associated with the spread of Chinese political influence throughout East Asia. In the late 2^{nd} century BC, the Chinese numerals were employed in tributary regions such as the Gansu corridor in central Asia, the Vietnamese states to the south, and the colony of Lelang (modern Pyongyang, North Korea), though of course the users of the script in all regions would have been primarily Chinese. The Chinese numerals were borrowed directly (without any transformation) by the Japanese as part of the *kanji* characters starting in about the 3rd century AD. The *hankul* alphabet developed in the 15th century for writing Korean has no corresponding numerical notation system, but the Chinese classical system was employed frequently in Korea. The numeral-signs associated with the *chu' nôm* script, which was developed in the state of Annam (in modern Vietnam), are simply the basic Chinese signs with additional phonetic notation; the basic Chinese system was also known and used in the region (Pihan 1860: 20-21). The numerals associated with the scripts of non-Chinese peoples of China, such as Tangut (Kychanov 1996) and Miao (Enwall 1994: 86) are also derived from the basic Chinese system, although sometimes with considerable modification. None of these systems is structurally distinct from the basic Chinese numerals, and thus describing them separately is not warranted. The Chinese numerals and their palaeographic variants enjoyed pride of place throughout East Asia until the introduction of Western numerals starting in the 17th century.

In addition, the Chinese numerals gave rise to three distinct descendant systems. Starting in the 10th century, China began several centuries of intensive contact with its neighbours to the north and west; warfare with these nomadic groups and the conquest of China in turn by the Kitan and Jurchin led to the development of Chinese-inspired numerical notation systems among these two groups.[6] Another structurally distinct descendant of the classical numerals is the so-called "commercial" system or Hangzhou numerals, which has been used since the 16th century in certain contexts related to commerce and trade. All three of these will be treated separately below.

---

[6] The Aramaic-derived scripts developed for writing the other languages of Central and East Asia (e.g. Uyghur, Mongolian, Manchu) do not possess distinct numerical notation systems, as far as I am aware.

For 1500 years after their inception in the Qin dynasty, the classical Chinese numerals were non-positional and used no zero-sign. Of course, the positional principle was known in China through the cumulative-positional rod-numerals that had been used since 400 BC. Moreover, Chinese mathematicians were aware of the use of ciphered-positional numerals in India and Southeast Asia in the 8th century AD. Qutan Xida[7], an Indo-Chinese Buddhist astronomer working at the Tang capital at Changan, first reported the use of nine unit-signs with a dot for zero in his great astronomical compendium, *Kaiyuan zhan jing*, which was written between 718 and 729 AD (Needham 1959: 12; Guitel 1975: 630-631). This transmission reflects the enormous scientific contact that accompanied the introduction of Buddhism into China in the 8th century AD. Yet the knowledge of ciphered-positional numerals had no impact on Chinese numeration for many centuries.

In the middle of the 13th century, a period of scientific vigour during the late Song dynasty, the first zero-signs appeared in Chinese mathematical and scientific texts. The first text in which a circle for zero was used with the classical numeral-signs is the *Shu shu jiu zhang* of 1247, the same document in which the zero-sign is first found with rod-numerals (Libbrecht 1973: 69). This modification allowed a circular zero-sign to be used whenever one of the decimal exponents in the middle of a numeral-phrase was empty. In theory, this would have allowed Chinese mathematicians to use only the unit-signs from 1-9 in conjunction with the 0 to express any number – thus transforming the system's structure from multiplicative-additive to ciphered-positional. Yet, during the Song dynasty zero was used only to fill in empty medial positions, while retaining the exponent-signs, so that where 12001 was written in the classical style as 一萬二千一, it is written as 一萬二千〇〇一 in the *Shu shu jiu zhang*, a form that is less concise than the classical one and provides no other obvious advantage. Only in the late 16th and early 17th centuries, when Chinese mathematicians of the Ming

[7] This name is the Sinicization of the author's original name, Gautama Siddharta.

dynasty were in extensive communication with the West, did the first ciphered-positional use of Chinese numerals occur. Tables of logarithms appeared at this time, using the nine basic unit-signs and a circle for zero in an identical way to the Western signs 0 through 9 (Menninger 1969: 461). Needham (1959) and Lam (1987) insist that we should regard positionality as having originated in China and spread to India (as I have already noted in discussing the origin of rod-numerals). The fact that zero was adopted so half-heartedly and in such an erratic way suggests that diffusion of the sign from Southeast Asia is much more likely.

Before the 16th century, zero was employed only in mathematical and scientific texts. After that point, it began to be used more widely, but rather than using the circular sign for zero found in the Song texts, a character, *ling* (零) 'raindrop', which had been used to designate remainders in division, began to be used in conjunction with the classical numerals. This sign was used in the sense of 'zero' throughout much of the Ming dynasty. The first text in which it featured prominently is the *Suan fa tong zong* of 1593, which is also the first text to describe the Chinese commercial numerals or *ma zi*, and additionally contains the first complete description of the bead-abacus or *suan pan* (Needham 1959: 16, 75-78). In this and other early texts, *ling* was used in exactly the same way as the circle-sign had been used previously, with one *ling* sign for every missing exponent, so that 30008 would be written as 三萬零零零七. While the introduction of the *ling* sign introduced an element of positionality into the system, it was not fully positional, since the exponent-signs were retained and *ling* was only used in medial positions. Chinese writers soon realized they could omit all but one *ling* when multiple consecutive exponents are empty, so that one could write 30008 simply as 三萬零七. The classical Chinese system normally uses *ling* in this manner today.

In modern China, any given number can be expressed in no less than six distinct ways, the choice of which depends greatly on context. Four of these forms are variants of the classical system. For literary and other prestige purposes, the pure classical Chinese numerals (without any sign for zero) are often used, thus representing a continuity of the signs and structure of the system from the Qin dynasty to the present. In most ordinary prose writing, some sign for zero is usually introduced in the medial positions, while retaining the exponent-signs. The use of *ling* has even spread to spoken Chinese, so that the preferred way to say 203 is not simply *er bai san* but rather *er bai ling san*. Where conciseness is desired or where it is desired to make each digit of a number clear (such as pagination or telephone numbers), the nine unit-signs along with a sign for zero are used in a ciphered-positional manner, as in the 17th century logarithm tables. In contexts where there is concern with forgery, the 'accountant's numerals' can be used. Another option is to use the commercial or Hangzhou numerals, which I will describe below; these are used mostly to record monetary values, and their use seems to be declining.

Table 8.10: Modern Chinese expressions for 20406

Classical	Classical with *ling* (zero)	Ciphered-positional	Western
二萬四百六	二萬零四百零六	二〇四〇六	20406

In most scientific and technical contexts in China today, Western numerals are preferred. Mao Zedong was certainly amenable (at least initially) to the use of Western numerals in place of Chinese ones, as indicated in a 1956 speech that was later suppressed (DeFrancis 1984: 262-263). Nevertheless, the replacement of Chinese with Western numerals has not been an uninterrupted or uncontested process. Some institutions reacted sharply to this trend, and anti-Western sentiment led to the replacement of Western numerals by the corresponding Chinese numerals in certain academic publications (DeFrancis 1984: 274-275). Western numerals are certainly well known to all reasonably educated people in China. In Japan and South Korea, the dominance of Western numerals is considerably greater than it is in China. Nevertheless, the Chinese numerals continue to be known and taught in these countries, though they are quickly acquiring an archaic flavour. In China itself, however, the use of local numerals shows no signs of sharp decline, and there is every reason to believe Chinese numeration will persist (at least in some of its forms) into the foreseeable future.

Chinese commercial

The Chinese commercial numerals (often known as 'Hangzhou numerals')[8] arose in the 16th century. The numeral-signs of the system are shown in Table 8.11 (Needham 1959: Table 22).

Table 8.11: Chinese commercial numerals

1	2	3	4	5	6	7	8	9
〡	〢	〣	〤	〥	〦	〧	〨	〩 or 久

10	100	1000	10000	0
〸	百 or ∋	千 or 丿	万	〇

Comparing these signs to those in Table 8.6, it can be seen that all of the unit-signs, save that for 5, are fairly clearly derived from the late forms of the rod-numerals used during the Ming dynasty, although they have been borrowed haphazardly from the *zong* and *heng* forms of that system. Because the unit-signs for 1, 2, and 3 use vertical rather than horizontal lines, they are derived from the rod-numerals rather than the classical system. Hopkins (1916: 318) explains the aberrant form of 5 as a form of the character *wu,* which is a homophone of the Mandarin numeral word for five. On the other hand, the most common versions of the exponent-signs for 10 through 10,000 are obvious variants of the classical system's exponent-signs. The circular sign for zero was in use in both the rod-numerals and the classical system. This evidence strongly suggests that the commercial numerals originated as a blend of the late rod-numerals and the Chinese classical numerals.

---

[8] Other names for this system include "ma zi", "Suzhou numerals", and "hua ma".

The system is multiplicative-additive in structure, with the zero used only to fill in empty medial positions, but never found at the end of numeral-phrases. Yet, unlike the regular Chinese numerals, where numeral-phrases are arranged in a single horizontal line, commercial numeral-phrases place the signs in two rows, with the unit multipliers of the various exponents on the top row and the exponent-signs, zero-signs, and the signs for the ones position on the bottom row (Pihan 1860: 6). Numeral-phrases were thus read in a zigzag fashion, starting at the top left, proceeding from top to bottom and then diagonally up and to the right.

This basic system was made more complex by a large number of irregularities, many of which were optional. When the number being expressed was a simple multiple of an exponent of 10 (e.g. 50, 800, 2000), the multiplier usually was placed to the left of the exponent-sign (as it would be in the classical system) rather than above it (Perny 1873: 101). When the number 10 occurred alone or in numbers such as 610 and 2010, the unit-sign 1 was always omitted, and the unit-sign could optionally be omitted when the sign for 10 was combined additively with unit-signs, as in numbers such as 18 and 212. Moreover, the special classical Chinese numeral-signs for 20 (廿) and 30 (卅) could be used in the commercial numerals where appropriate (Hopkins 1916: 319). When there were two consecutive zero-signs in a numeral-phrase, they could be placed one atop the other rather than side by side in the bottom row, as would be normal. Finally, the standard classical unit-signs for 1 through 3 (horizontal rather than vertical strokes) are sometimes used in the units position at the end of numeral-phrases, though they cannot be used as multipliers in conjunction with exponent-signs (Hopkins 1916: 319). The combination of all these irregularities and options means that almost any number may be expressed in several valid ways. Table 8.12 depicts a selection of numeral-phrases as written in this system.

Table 8.12: Chinese commercial numeral-phrases

40709	〤 〧 万〇百〇〩	4 7 10,000 0 100 0 9
26	〢 十〦 OR 廿〦	2 10 6 OR 20 6
162	〡〦 〡〦 百十〢 OR 百十二	1 6 100 10 2
917	〩〡 〩 百十〧 OR 百十〧	9 1 9 100 10 7 OR 100 10 7
3008	〣〇 千〇〨	3 0 1000 0 8
5000	〥千	5 1000

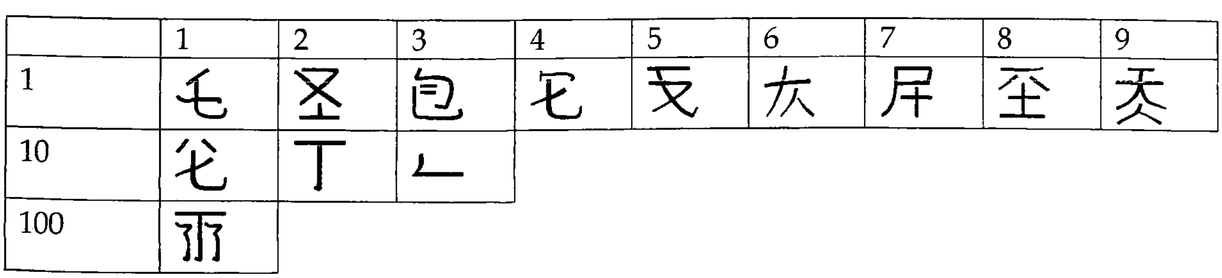
We do not know exactly when the commercial numerals were invented, but the earliest printed text that describes them is the *Suan fa tong zong*, published in 1593 (Needham 1959: 5). Because they were not used for prestige purposes, such as literature or mathematics, but were restricted to a limited set of commercial contexts (invoices, bills, signs for prices, and so on), earlier evidence of their use may not have survived. It is safe to assume that they were used throughout the 16th century. The rod-numerals, from which the commercial numerals are partly derived, were obsolescent by 1600. It thus seems unlikely that they would have been used as the basis for a new system as late as 1593. Yet early texts that mention them associate their invention and use with the great commercial city of Suzhou (in Jiangsu province). As this city only came to prominence in the 16th century, if the attribution of their invention to Suzhou is correct, a pre-16th century origin seems unlikely.

As is suggested by their name, the commercial numerals were (and are) used solely in commercial contexts. They continue to be used even today in some parts of China on bills, invoices, and signs in shops and markets (primarily to indicate prices), though their use appears to be waning in favour of regular Chinese numerals or Western numerals. They seem to be most common in regions where Cantonese is spoken, including Hong Kong.

Kitan

The Kitan (or Khitan) were an Altaic-speaking people who ruled Manchuria and other parts of northern China between 916 and 1125 AD (now known as the Liao Dynasty by the Chinese). While there was no Kitan writing before their conquest of Manchuria, two scripts were developed shortly thereafter, the 'large script' and the 'small script', both based largely on the Chinese script. Neither of these scripts is fully deciphered, because the Kitan language is only poorly known, but the meanings of the Kitan numeral-signs are understood. The numerals of the 'large script' are identical to the classical Chinese numerals used during the Song dynasty that ruled southern China at the time of the Kitan conquest. The 'small script', purportedly developed by the Kitan scholar Diela during the visit of an Uyghur delegation to the Kitan court in 924 or 925 AD, uses a set of numerals that are quite distinct from the Chinese system. Despite this impetus, it was clearly the Chinese rather than the Uyghurs to whom the Kitan looked for a model for their script and numerals. The signs of this system are shown in Table 8.13 (Kara 1996: 233).

Table 8.13: Kitan numerals



While the Kitan numeral-signs have a vaguely Siniform appearance, they are entirely dissimilar to the corresponding Chinese numerals, and may be presumed to be of indigenous origin. Numeral-phrases are multiplicative-additive and are read vertically from top to bottom. A slight ciphered element is introduced into the system in the existence of distinct characters for 20 and 30; this practice is probably derived from the analogous Chinese signs, 廿 and 卅, although the Kitan signs are not cumulative. It is not known how (if at all) numbers higher than 1000 were written. In the Kitan numerals, 473 might be written as follows:

473 =

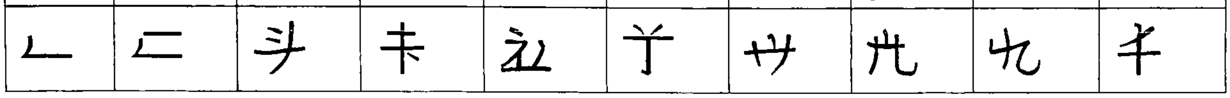

Because Kitan writing is so poorly understood, it is difficult to know the total scope of contexts in which the numerals were used; most texts were probably historical records of events, in which numerals are used primarily for dating. The Kitan script and numerals did not long outlast the period of Kitan independence, which ended in 1125 at the hands of the Jurchin. In 1191, the use of the Kitan script was forbidden by Chinese imperial order, after which time no further instances of its use are attested (Kara 1996: 231).

Jurchin

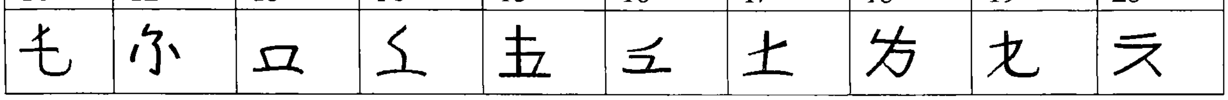
The Jurchin (also Jurchi or Jurchen) were the rulers of what is now known as the Jin Dynasty in the northern part of China (1115-1234). Soon after establishing their dynasty, the Jurchin developed their own script (a mixture of ideograms and syllabograms) to which was attached a set of ideographic numerals. These numerals are shown in Table 14 (Grube 1896: 34-35).

Table 8.14: Jurchin numerals

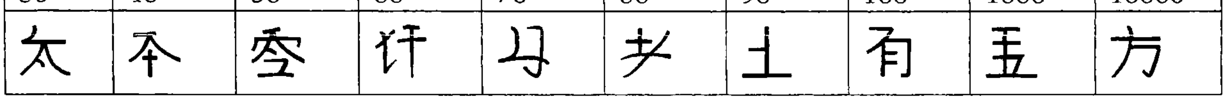
1	2	3	4	5	6	7	8	9	10



11	12	13	14	15	16	17	18	19	20



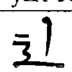
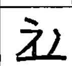
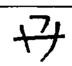
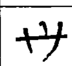
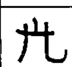
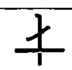
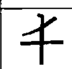
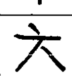
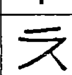
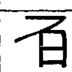
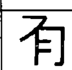
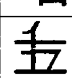
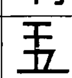
30	40	50	60	70	80	90	100	1000	10000



The Jurchin numerals are primarily decimal, although they contain traces of a vigesimal system in that there are distinct numeral-signs for 11-19, none of which can be derived from additive combinations of 10 and the appropriate units. For writing numbers from 20 to 99, unit-signs from 1 through 9 sometimes were combined with the exponent-sign for 10 as in the classical Chinese system, which means that the Jurchin numerals appear to be multiplicative-additive. Yet there were Jurchin numeral-signs for 20 through 90 that were used in a ciphered rather than a multiplicative fashion. For numbers above 100, the multiplicative principle was always employed. Thus, the Jurchin system is structurally closer to hybrid ciphered-additive / multiplicative-additive systems, such as the Ethiopic numerals (ch. 5) and Sinhalese numerals (ch. 6), than it is to Chinese. In the Sino-Jurchin texts from the Ming Dynasty published by Grube (1896), which date roughly to the period 1450-1525, only the unit-signs 1-9 and the exponent signs 10, 100, 1000 and 10,000 were used. It is not clear with what frequency or in what contexts the Jurchin used the signs for 11-19 and 20-90. Regardless of period, Jurchin numerals, like the script, were written in vertical columns read from top to bottom, with the highest-valued exponents at the top.

The origin and early history of the Jurchin numerals is adequately documented (Kara 1996: 235). A Jurchin 'large script' was introduced in 1120 by Wanyan Xiyin, and was based on the Kitan script with significant Chinese influences; the script was officially introduced in 1145 by Emperor Xizong, with a number of 'small script' characters added. The Jurchin numerals are found on many monuments of the Jin Dynasty and some manuscript fragments. The writings that survive are historical and literary in nature, and the numerals on them are mainly dates. Our best evidence for them comes from the Ming Dynasty (1368 – 1644), when Chinese translators produced a bilingual glossary and translated documents (Kara 1996: 235). It is from this glossary that the numeral-signs above are taken; Kara (1996: 236) provides a less complete (yet structurally identical) set of earlier signs. Some of these differences are indicated in Table 8.15.

Table 8.15: Early and late Jurchin numerals

	Jin form (Kara 1996)	Ming form (Grube 1896)
5		
7		
8		
10		
20		
100		
1000		

Although the Jurchin did not control large regions of China for very long, the Jurchin script survived for several centuries. It was used on a Ming inscription of 1413, suggesting that it was not simply a historical curiosity, but was being preserved because it was being used (at least by some people). While it was certainly obsolescent after the middle of the 14th century, it continued to be used until at least 1525, at which time Ming translators were still working with Jurchin documents. The Jurchin were one of the major constituent groups of the Manchu who conquered China in the 17th century (in fact, the ethnonyms 'Jurchin' and 'Manchu' may refer to a single group), but by this time they used either the classical Chinese numerals or the ciphered-positional, Indian-derived Mongolian numerals.

Ryukyu

While the meagrely populated Ryukyu Islands seem an unlikely locus for numerical creativity, three different numerical notation systems have their origin in this tiny Pacific archipelago south of Japan. The first of these, and certainly the least interesting from our perspective, are a set of numeral-signs from 1 to 10, which are no more than slight paleographic variants of the traditional Chinese numerals (Pihan 1860: 18-19).[9] While we do not know how numbers higher than 10 were formed using this system, it is probable, given the similarity of these numeral-signs to those of China and Japan, that it was a multiplicative-additive system. The second system was a form of knot-notation known as *ketsujo*, by which amounts of money were counted using series of knotted ropes that were strung perpendicular to a long cord, in a way that is analogous to the Peruvian *quipu* (Ifrah 1998: 543). This system roughly corresponds with a cumulative-positional numerical notation system with a base of 10 and a sub-base of 5. Unfortunately, too little evidence is available to analyze the *ketsujo* system in detail.

The third system was written on long wooden sticks (30 to 75 cm in length, and 2.5 to 4 cm in breadth), which were known in Okinawan as *sho-chu-ma* (Chamberlain 1898). It comprises several variants, each of which was used for a particular commodity: money, bundles of firewood, bags of rice, and possibly other goods as well. While these sticks have been described as 'tallies', the marks do not count objects in sequence (one mark for one object), but constitute a full-fledged numerical notation system used for

[9] While the Ryukyu Islands have been under Japanese control since the 17th century, the cultural influences in the archipelago have been at least as much Chinese as Japanese, given its location in the East China Sea.

recording amounts of goods. Unfortunately, this system is known only through Chamberlain's paper of 1898, so we are lacking much useful information about it. Tables 8.16 and 8.17 show two of the more common systems used in the late 19th century, the first for expressing quantities of money (in units of *kwang* and *mung)* and the second for counting bundles of firewood (Chamberlain 1898: 385, 388).[10]

Table 8.16: Ryukyu numerals (money)

	1	2	3	4	5	6	7	8	9
10 mung					⺈				
100 mung	•	••	•••	••••	┐	•┐	••┐	⋱┐	⁘┐
1 kwang	—	=	≡	☰	∟	∟ —	∟ =	⊤ ≡	⊤ ☰
10 kwang	+	干	丰	丰	◡	𐊈	𐊈	𐊈	
100 kwang	○ ⊙	⊙	⦶ ⊙		⊤◡				
1000 kwang	⊕				⊤◡				
10000 kwang	◎								

Table 8.17: Ryukyu numerals (firewood bundles)

	1	2	3	4	5	6	7	8	9
1	•	••	•••	••••	∟		∟ ••		
10	+	⊥	土	丰	Y	Y	Y	Y	Y
100	○	⦶	⦶	⊕	◡	○	⦶	⦶	⊕
1000	⊙								

The signs shown in the tables reflect those attested on the *sho-chu-ma* examined by Chamberlain; nevertheless, the form of most of the non-attested signs can be easily inferred on structural grounds. While the numeral-signs are slightly different in these two systems, both are cumulative-additive and decimal, with a sub-base of 5. The multiples of each exponent from 1 to 4 are mainly cumulative (exceptions include the

[10] I have corrected a couple of errors in Chamberlain's tables where numeral-signs were clearly assigned incorrect values.

"100 kwang" money count and the hundreds value in the firewood count), and the multiples from 5 through 9 are expressed by combining the appropriate sign for 5 with the required number of additional units. The numeral-signs are largely abstract. In some cases, the sign for 5 of a given exponent is derived by halving the sign for one of the next higher exponent, such as  for 500 and  for 1000 in the firewood system or  for 5000 kwang and  for 10000 kwang in the money system. Numeral-phrases are written in a roughly vertical fashion, such as , indicating 352 kwang and 250 mung.

While the numerals used on the *sho-chu-ma* are simple enough to understand and use, their origin is obscure. We do not know when they first began to be employed, or under what circumstances. Chamberlain states, "The custom may be traced to a hearsay knowledge of the Chinese written character among the Luchuan [Ryukyu] peasantry, who, not possessing sufficient learning to employ this character itself, and not being encouraged by their rulers to acquire the elements of an education deemed unsuitable to their lowly station, developed a make-shift of their own" (1898: 383). Indeed, it is probable that the sign for 10 bundles / 10 kwang, 十, was borrowed directly from the identical Chinese sign for 10 (Chamberlain 1898: 384). If Chamberlain is correct, the Ryukyu system was produced by stimulus diffusion rather than direct diffusion from the Chinese classical numerals. Since the Japanese also used the Chinese numerals, the Okinawans may have learned the system from Japan rather than China. Moreover, a certain similarity between the "1 kwang" signs and the rod-numerals may be detected, and furthermore, the rod-numerals are cumulative (though positional) and quinary-decimal, so it may be possible that the rod-numerals were ancestral to the Ryukyu tallies. Yet the rod-numerals had fallen out of use for most purposes by the 17th century and the idea of using lines for units is nearly panhuman. We require more evidence on the history and early numeral-signs of the Ryukyu system in order to confirm this idea.

By the time these numerals were reported in the Western scholarly literature at the end of the nineteenth century, the Ryukyu numerals had already ceased to be used, having become a historical curiosity, or even an object of embarrassment, for the Ryukyuans (Chamberlain 1898: 383). It is indeed unfortunate that no more information can be obtained about this system, because it is the only cumulative-additive system ever used in East Asia. Moreover, its sub-base of 5 and the use of halvings of the main exponent signs for the fives render the Ryukyu system structurally identical to the Etruscan and Republican Roman numerals (ch. 4), which also have their origin in tally-style marks and make use of the principle of halving.[11] We may never be able to learn more about the development of a remarkable and heretofore unacknowledged parallel invention.

Summary

Chinese numerals are central to the history of the East Asian family. Today, the classical Chinese numerals (along with positional variants) occupy a role parallel to the supremacy of the Roman numerals in Europe prior to 1500, despite the increasing use of Western numerals for science, technology, and commerce. While it is perhaps inappropriate to speculate on the future of the Chinese system, its continued strength (at least in China) suggests that it will continue to thrive, especially in non-technical prose writing. We must also take into account the strong cultural preference for Chinese symbol systems when analysing the present state of the Chinese numerals; functional considerations alone cannot account for it. The increasing rarity of the Chinese numerals in Japan and Korea is probably not simply the functional rejection of an "inefficient" system, but rather an active resistance against a Chinese cultural feature in favour of the more international Western numerals.

As described above, the Chinese numerical notation system as used today is enormously variable in structure, and employs a host of representational techniques. On

[11] Of course, it is highly improbable that the Roman numerals are an ancestor of the Ryukyu system.

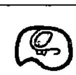
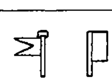
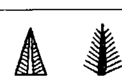
the surface, this appears hopelessly non-functional, and we might question why such a system would survive. I think that its quasi-lexical nature - the fact that Chinese numerals act as both ideographic script-signs and semasiographic numeral-signs - renders this variability both comprehensible and rational. If Western numerals had to incorporate archaisms such as *score*, or account for the fact that 1400 can be *one thousand four hundred* but is more commonly *fourteen hundred*, no doubt the same eccentricities would occur in our own numerical notation system. The Chinese classical numerals are well suited to being read because they account for the irregularities in spoken Mandarin. Moreover, the basic multiplicative-additive structure of the system permits all sorts of structural manipulations, such as the occasional use of positionality or the use of ciphered signs for the lower decades, without any ambiguity arising in numeral-phrases. The system's flexibility and its correspondence with speech are thus advantages rather than hindrances.

The comparison of this phylogeny with the ones I have discussed previously is quite instructive. In Chapters 2 through 7, most systems of each family employed a single common structural principle. In contrast, the systems of the East Asian family display all five major structural principles - cumulative-additive (Ryukyu), cumulative-positional (rod-numerals), ciphered-additive (Jurchin), ciphered-positional (Chinese positional variant) and, of course, multiplicative-additive (Shang/Zhou, Chinese classical and commercial, Kitan). Yet there can hardly be any doubt that these systems comprise a cultural phylogeny. The historical connections among systems are well established, and the similarities in the numeral-signs are quite strong. Only the earliest system, the Shang, is a completely independent invention. If we were to rely on structural qualities alone, we would be at a loss to describe the cultural history of this family.

Chapter 9: Mesoamerican Systems

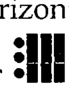
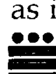
All the numerical notation systems I have discussed so far have been Old World inventions. Yet many New World civilizations also used written numeration. In this chapter, I will discuss the numerals used in Mesoamerica between 400 BC and 1600 AD: those of the Mayan and highland Mexican (Aztec) civilizations prior to the Western conquest of the Americas. I will treat other unrelated New World inventions, such as the Inka *quipu* and the Cherokee numerals, in Chapter 10. In past research, the primary theoretical importance of the Mesoamerican numerals has been to provide clear New World examples of independent invention of features of numerical notation systems such as additive notation (Guitel 1958) and the zero (Kroeber 1948: 468-472; but cf. Seidenberg 1986). Because the cultural history of the numerals of Mesoamerica is poorly understood, however, we do not know with certainty how the lowland and highland Mesoamerican systems are related to one another. Paradoxically, Mesoamerican numerals are rarely studied today because, along with calendrical signs, they were the earliest aspect of the region's representational systems to be deciphered and thus are among the best understood. Yet to claim that the Mesoamerican systems are fully understood would be an exaggeration. In fact, serious misinterpretations of the data continue to plague our understanding of this family. The numeral-signs of its major systems are shown in Table 9.1.

Table 9.1: Mesoamerican numerical notation systems

System	1	5	20	400	8000	0
Bar and dot (stelae)	●	|				
Bar and dot (codices)	●	—				
Aztec	●					
Texcocan line and dot	|		●			

Bar and dot

The bar and dot numerals were the most commonly used system in Mesoamerica. While most often encountered in the texts of the Maya civilization, bar and dot numerals were ubiquitous in all the scripts of lowland Mesoamerica, from stone monuments (400 BC - 910 AD) to the four surviving Maya bark-paper codices (1000 - 1500 AD). This system has been the object of study for over a century (Bowditch 1910; Morley 1915). It was the first Maya representational system to be deciphered, and its interpretation is thought to be very secure. Its frequency of use reflects not only the strong lowland Mesoamerican interest in dating and calendrics, but also the practice of incorporating numerical values into the names of Maya deities. Yet the bar and dot numerals are very simple in structure and use only a handful of numeral-signs, of which the dot for 1 and a bar for 5 are the most common, hence the system's name.

The numbers from 1 to 19 are written by combining the dot sign for 1 and the bar sign for 5 additively. When the bars are vertical, as is most common on stone inscriptions, they are usually placed to the right of the dots, but they are placed below the dots when the bars are horizontal, as in the codices and a few monumental texts. Thus, 18 can be written as either  or . Short numeral-phrases such as these were most often combined with another glyph indicating the thing being quantified. Mesoamerican hieroglyphic writing on stone was a very ornate art, and numerals were often altered or ornamented in various ways that can make reading a numerical value difficult. Ornamental crescents were often employed in order to 'fill in' a numeral that would otherwise have an empty space, and these can easily be confused with dots; thus  means 11 rather than 13 (Thompson 1971: 130). Similarly, decorative lines were sometimes added to bars for aesthetic purposes, which makes it difficult to distinguish one from two bars on some inscriptions.

In addition to the bars and dots of this system, a sign for 20 was also occasionally used. Many historians of mathematics have neglected the sign for 20 entirely, in part because it is comparatively rare in Mayan inscriptions. However, the use of this sign, combined with the bar and dot numeral-phrases, produces a base-20 cumulative-additive system with a sub-base of 5. Whether it occurs in the Maya codices, as , or on stone inscriptions, as , it has the numerical value of twenty (*kal*), but it is also a glyph meaning 'moon' or 'lunar month' (Lounsbury 1978: 764).[1] A related sign, , also meaning 20 or moon, was used in epi-Olmec inscriptions such as the La Mojarra stela, which dates to 159 AD (Justeson and Kaufman 1993). It can occur on its own or in conjunction with bar and dot numerals from 1 to 19, thus representing numbers as high as 39. However, it is never repeated in a numeral-phrase (that is, one would not write 60 as ). Kelley (1976: 23) lists many examples of numeral-phrases using this glyph which show that the accompanying bar and dot numerals could be placed above, below, or to either side of a 20-glyph. Thompson (1971: 139) indicates that the 20-sign was only used to indicate intervals between dates that were greater than 20 but less than 39 days, thus avoiding the use of combinations of *uinals* (periods of 20 days) and *kins* (1 day). Very rarely, it was used in expressions for larger time intervals, such as 2 *tuns* (periods of 360 days) and 36 days found on Stela 22 from Tikal (Closs 1986: 344). It is also found in an irregularly constructed date on Stela 5 at Pixoy, indicating a quantity of 20 tuns (periods of 360 days) (Closs 1978). In a few instances, the 20-glyph was used for non-calendrical counts as well; these unusual examples will be discussed below.

A glyph that essentially means "zero" was also used in the bar and dot numerals. There is considerable paleographic variation in the signs used for zero, but a 'shell' sign,  or , was commonly used in the codices, while different signs, such as  and

[1] Closs (1978: 691) notes that the central dot in the latter of these signs is only found on inscriptions where the glyph has the numerical value '20', thus distinguishing it from the more generic 'moon', where the dot is missing.

, were used in monumental writing. A significant debate regarding the meaning and function of the zero-sign concerns whether we should interpret it as 'zero' or 'completion' (cf. Thompson 1971: 137). The 'completion' position suggests that the Maya zero is quite different from the Western one, and that the glyph should be interpreted as a placeholder with the rough meaning of 'completion of a given cycle of time'. This interpretation reflects the function of the sign, which normally is used as the numerical coefficient of glyphs for different periods of time. Yet I do not see any reason to deny the Maya their zero. The Maya zero-sign is clearly numerical in function; it is found in the same contexts as the regular bar and dot numerals, and so the meaning 'zero' is more appropriate than 'completion'. While the Maya probably did not have an abstract concept of zero, as is present in Western mathematics, neither did the Babylonian astronomers (ch. 7), for whom the zero-sign served as the marker of an empty medial position but not as an abstract number. Even if the Maya zero-sign does not represent an abstract concept of zero, this does not diminish its importance for the system's structure.

While bar and dot numeration is most closely associated with Mayan civilization, the bar and dot numerals existed in the very earliest Mesoamerican scripts, starting with the Zapotec inscriptions of the latter part of the Middle Formative period (ca. 500 - 400 BC). The earliest of these is Monument 3 from San José Mogote in the Valley of Oaxaca, where the day-name "1 Earthquake" is written with a stylized dot (Marcus 1976: 44-45). Stela 12 from Monte Alban provides the first example of a combined bar and dot phrase,  (8), apparently indicating a day of the Zapotec month (Marcus 1976: 45-46). Colville (1985: 796) is agnostic as to whether the bar and dot system was invented by Mixe-Zoqueans (such as the Olmec) or the Zapotec, since both used vigesimal lexical numerals with a quinary component. That the Middle Formative bar and dot numerals are all from Oaxaca suggests that they were a Zapotec rather than an Olmec invention. Regardless of whether the Zapotecs or late Olmecs first developed them, though, they were clearly an independent Mesoamerican invention. Hyperdiffusionistic arguments, such as those of

Seidenberg (1986), which assume that multiple instances of independent invention are nearly impossible, cannot deal adequately with the early date of these inscriptions, the unique quinary-vigesimal structure of the system, or the general increase in the frequency and complexity of bar and dot numeral expressions over time.

In the Late Formative period (400 BC to 50 AD), we continue to find bar and dot numerals on Zapotec inscriptions in Oaxaca, but others appear on the Gulf Coast in the inscriptions of the latter stages of the Olmec civilization (the so-called 'epi-Olmec' script). In the 1st century BC, we find the first examples of bar and dot numerals arranged in vertical columns indicating periods of time, such as on Stela 2 from Chiapa de Corzo, dating to 36 BC, and Stela C from Tres Zapotes, dating to 31 BC (Marcus 1976: 49-53). The longer and somewhat later epi-Olmec inscriptions, such as the La Mojarra stela (159 AD) and the Tuxtla statuette (162 AD), both contain many bar and dot numeral-phrases (Justeson and Kaufman 1993). The first securely dated *Maya* inscription that uses this notation is Stela 29 from Tikal, which dates to 292 AD (Lounsbury 1978: 809); however, Stela 5 from Abaj Takalik, which dates to 126 AD, may also be an early Maya inscription (Closs 1986: 327). These early bar and dot numerals are written with dots placed above horizontal bars and numerals were not attached to any other glyph (although in these cases, we can tell that the quantities enumerated were periods of time). This means of representation would later be the standard practice in the Maya codices; however, most later Maya monumental numerals were written with vertical bars with dots to their left. It is unclear what prompted this change in the technique of representation.

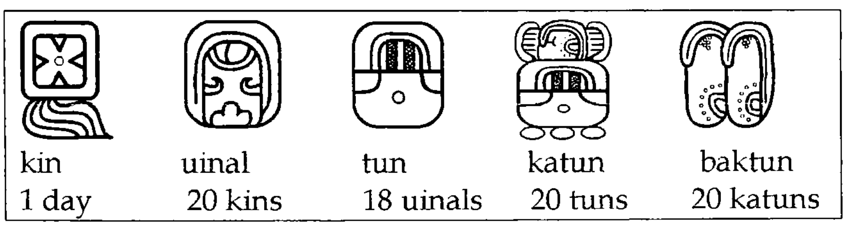
In the Maya Classic period (250 - 900 AD), the bar and dot numerals are ubiquitous on stone inscriptions. There is no identifiable regional variation in the form or ornamentation of the numerals within the Maya sphere of influence. Very early in the Classic period, the bar and dot numerals spread through highland Mexico; the Mixtecs used bar and dot notation to write numbers up to 13 (Caso 1965: 955). It is not clear whether they borrowed them from the Zapotecs (who also lived in the Oaxaca region) or from the more influential Maya. Bar and dot numerals were used occasionally at Teotihuacán, where there was no indigenous script that represented phonetic values. Langley (1986: 139-142) notes about a dozen secure instances where numbers less than 13 were written with bars and dots, and a few other less likely examples. There is some evidence that the bar and dot numerals survived in central Mexico until the Spanish conquest. In Mixtecan-Pueblan texts such as Codex Fejervary-Mayer and Codex Cospi, sets of bars and dots arranged vertically or horizontally could represent counted bundles of offerings (Love 1994: 61). In addition, although it is an irregular formation by the normal rules of the system, a set of four bars from the Postclassic Codex Selden, which was written in the Mixtec manuscript tradition, may represent a quantity of twenty bundles (Boone 2000: 43). However, to my knowledge, the Aztecs never used bar and dot numerals, instead relying on their own (purely vigesimal) cumulative-additive numerals.

After the collapse of classic Maya civilization in the 10th century AD, attested examples of bar and dot numerals became increasingly rare. The latest Maya monumental inscription dates to 909 AD (Closs 1986: 317). Many regions where bar and dot numerals had previously been used, such as Oaxaca and the Valley of Mexico, abandoned the old system in favour of the central Mexican dot-numerals, which represented numbers from 1 to 19 through dots alone, and did not represent numbers higher than 19 at all. Bar and dot numerals were retained during the Postclassic period (10th to 16th centuries) in Guatemala and Yucatan, where they were used on bark-paper codices until the Spanish conquest (Urcid Serrano 2001: 3). The last text on which bar and dot numerals occur is one of the books of Chilam Balam, in which an annotated description of the system is dated 1793 (Thompson 1971: 130). Yet the system essentially had ceased to be used by 1600 and was replaced by Roman or Western numerals for all purposes.

Maya calendrics and 'positional' bar and dot numeration

The most common recorded function of bar and dot numerals was counting periods of time. A bar and dot numeral-phrase from 0 to 19 would often be combined with one of five glyphs for time periods - *kin* (1 day), *uinal* (1 'month' of 20 *kins*), *tun* (1 'year' of 18 *uinals*), *katun* (20 *tuns*), and *baktun* (20 *katuns*) – by placing the numeral to the left of the period-glyph.[2] Each successive period is twenty times the previous one, except for the tun of 18 uinals, which comprises a sum of 360 days in order to correspond roughly with the calendar year.[3] Some of the more commonly used signs for these periods are shown in Figure 9.1 (cf. Closs 1986: 304-5)[4]:

Figure 9.1: Maya period glyphs



To express a specific fixed date, five numeral-glyph combinations were required (one for each period, written from longest to shortest). These were normally written in pairs of columns, reading from left to right and top to bottom, although other directions of reading are not unknown. These were what we now know as the "Long Count" dates of Mayan civilization, expressing the amount of time between the starting point of the Maya calendar (corresponding to the date August 10, 3113 BC in the widely accepted

---

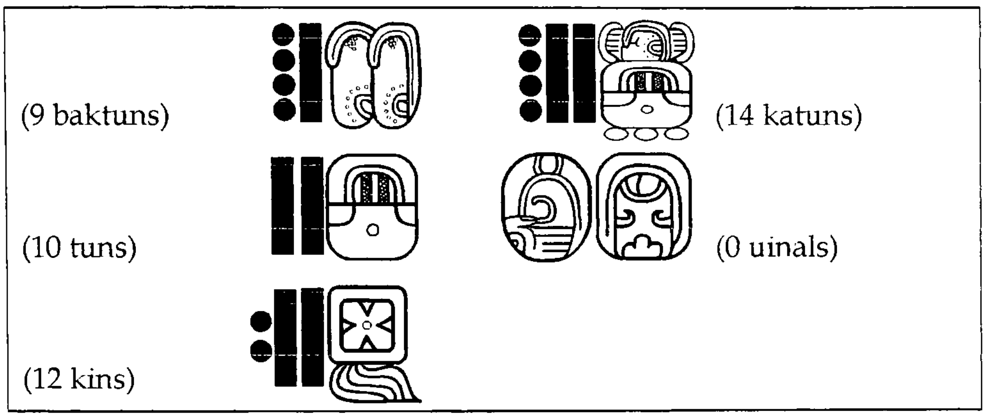
[2] The terms *katun* and *baktun* mean, literally, '20 tuns' and '400 tuns'. The latter term is in fact a coinage of Mayanists; there is no evidence that this word was associated with the glyph in question in ancient times. There are several extremely rare glyphs for longer periods, again with coined names: *pictun* (8000 tuns), *calabtun* (160,000 tuns), and *kinchiltun* (3,200,000 tuns), which presume a purely vigesimal progression of dates (Closs 1986: 303).

[3] It appears, however, that the Yucatecan and Cakchiquel Maya may have had a purely vigesimal year of 20 months of 20 days, though their *numerical notation* does not reflect this fact. (Satterthwaite 1947: 8-9)

[4] I have intentionally omitted the enormous amount of paleographic variability among these signs, because they are non-numerical.

Goodman-Martinez-Thompson correlation with the Gregorian calendar) and any other date. In addition, the amount of time between any two days could be expressed by a 'Distance Number', such as 12 tuns, 0 uinals, 4 kins. Mayanists use a convention whereby time values are expressed by writing the five numerical coefficients separated by points; thus, the Long Count date shown in Figure 9.2 would be written as 9.14.10.0.12 by modern scholars.

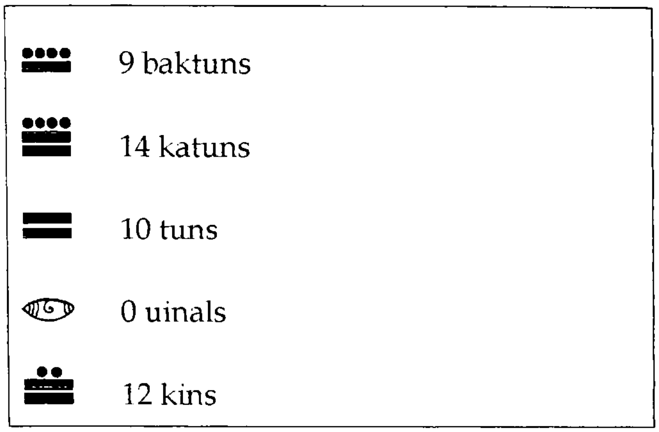
Figure 9.2: Long Count date (9.14.10.0.2) with period glyphs



For both Long Count dates and Distance Numbers, if the coefficient of a given time period was zero, the Maya would include both a zero coefficient and a period glyph for that value, even though it was not logically necessary to do so in order to interpret the phrase correctly. Thus, in Figure 9.2, the value "0 uinals" could have been omitted without any loss of meaning, but it was normally included (Closs 1986: 306-7). While it is not known exactly why the Maya did this, it was probably for aesthetic reasons. Very occasionally, in Distance Numbers (though never to my knowledge in Long Count dates), a period with a coefficient of zero was suppressed entirely (Thompson 1971: 139).

This question of the suppression of glyphs seems rather trivial at first, but it is crucial if I am to rectify a major misconception about the nature of bar and dot numerals. In a few classical monumental texts and in the Dresden Codex from the Postclassic period, period-glyphs were omitted entirely, and dates were written simply by placing the five coefficients in a single vertical column, using the vertical bar and dot numerals and the 'shell' sign for zero. Figure 9.3 shows the Long Count date 9.14.10.0.12 as it would be written in this manner.

Figure 9.3: Long Count date without period glyphs



This system requires that all the relevant numerical coefficients be included, even for periods for which there is a zero coefficient, to ensure that the correct quantity of time is counted. The bottom value always represents kins, the second from the bottom uinals, and so on, preventing any misreadings. Because these units of time are arranged in a mainly vigesimal sequence - each higher value is equal to twenty of the next lower value, save that 1 tun is equal to 18 rather than 20 uinals - the similarity between this system of writing dates and a base-20 cumulative-positional numerical notation system (with a sub-base of 5) is striking. The consensus among Mayanists today is that, in fact, this system of dating represents a positional numerical notation system with a zero.

For this to be true, we must presume that, when the Maya wrote number columns such as the one in Figure 3, each position represented a particular component of a single number. Positional numerical notation systems do this by having each successive position represent the next higher exponent of a given base. Thus, when I write the number 1942, I mean a single count of some quantity, of which there are 1942, consisting of one thousand, nine hundreds, four tens, and two ones. In the Maya case, where the lowest unit expressed is *kins*, it is quite natural to assume that if there were a positional numeral system, it would count *kins*. It is quite simple to translate the five time periods into counts of days and then to take the sum, as shown in Figure 9.4.

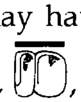
Figure 9.4: Positional Maya count of days

9 x 144,000 days	1,296,000 days
14 x 7,200 days	100,800 days
10 x 360 days	3,600 days
0 x 20 days	0 days
12 x 1 day	12 days
	= 1,400,412 days

Most Mayanists implicitly assume that these vertical columns of bar and dot numerals, without period glyphs, are to be understood as an integer representing a count of days (Kelley 1976; Marcus 1976; Lounsbury 1978). If this is in fact the case, then we have a *bona fide* cumulative-positional numerical notation system. Yet, if the period-glyphs were meant to be inferred when reading these columns, then such numerals can be read as five separate values, just as they would be if the glyphs were included. How, then, can we tell whether the interpretation in Figure 4 is one that the Maya themselves made, or whether they simply 'read in' the missing period-glyphs? Is the correct interpretation "1,400,412 kins" or "9 baktuns, 14 katuns, 10 tuns, 0 uinals, 12 kins"? For a number of reasons, I believe that the latter interpretation is more likely. No matter how much these columns of numbers may look as if they are a positional means of representing a single number, I think that Maya dates were read, with or without period-glyphs, in an identical fashion. I therefore consider dates written in vertical columns without period-glyphs to be a quasi-positional calendrical system rather than a fully positional system for representing large numbers.

A neglected tradition in the study of Maya calendrics and numeration recognizes that Long Count dates (with or without period-glyphs) are capable of being read positionally or non-positionally. While most Mayanists assume that the Long Count is a simple count of days, a small and often-overlooked body of Mayanist research over the past 75 years has tried to deal with the problematic interpretation of the Maya Long Count. Teeple (1931) and Thompson (1971) claimed that the Long Count dates should be considered as a count of tuns (years), in which the final two places (uinals and kins) represented two separate fractions of years. Satterthwaite (1947) held that they should be read as two separate counts, one of years (the first three positions), the other of days (the last two). Closs (1977), the most recent scholar to deal seriously with this issue, felt that there were in fact three counts: a tun count, comprising, first, a positional numeral indicating 1, 20, and 400 tuns, second, a non-positional bar and dot numeral indicating uinals, and third, a non-positional bar and dot numeral indicating kins. All agree that the highest three periods (baktuns, katuns, and tuns) were read and understood by the Maya as a single count of tuns. Moreover, they claim that the Long Counts were understood in this way, *whether or not the period-glyphs were present.* These readings are made on the basis of several lines of evidence. Separating the higher values, which are purely vigesimal, and the lower ones, in which the 18 uinals = 1 tun irregularity occurs, renders the system more readable, given the purely vigesimal structure of the Maya lexical numerals. It also helps explain a number of texts where the glyphs for the tun and its multiples are distinguished (by colour or ornamentation) from the other two (Closs 1977: 22-23). I agree fully with Closs that the kin, uinal, and tun counts were read separately, but believe that he has not gone far enough. There is no reason to think that the Maya wrote glyphs for the baktun and katun but then simply ignored them in reading, instead multiplying out a sum of years. I thus regard the Maya Long Count as five separate non-positional counts of five different time periods.

If bar and dot numerals were used for large quantities of things other than time, this discussion would be moot, because there would be clear instances where the higher positions represent *exponents of a base*, rather than long calendrical periods. Yet Mayanists and scholars of numeration alike have failed to emphasize that the "positional" numerals are not used for just any sort of quantity, but only for counts of time periods. There are *no* Mesoamerican texts where "positional" bar and dot numerals were used to count quantities of goods, numbers of people, or anything but amounts of time. While there is ethnohistorical evidence from Yucatan suggesting that some form of written numeration was used by the Mayan traders and administrators, it is overextrapolation to postulate, as Lounsbury does, that positional bar and dot numerals were used throughout Maya history for trade, tribute, mensuration, and other functions (Lounsbury 1978: 764).

When the Maya wrote larger numbers of quantities other than time, they often used bar and dot numerals, but never the sort of vertical columns described above. Very rarely, they used additive techniques, such as the moon-glyph for 20, which is used in counts of 20 and 21 captives (S. Houston, personal communication). In other cases, it is possible that multiplicative techniques were used. On several pages of the Paris Codex, long series of numbers between 1 and 19 are followed by a 20-glyph. Love (1994: 57-59) has interpreted these as a sort of multiplicative formation by which each of the bar and dot signs was multiplied in value by 20 in order to represent scores of ritual offerings. As well, bar and dot numerals could be combined with basically non-numerical signs. On a mural from Bonampak, it has been suggested that a bar numeral for 5 was combined with a glyph, *pi*, which may have stood for a unit of 8000 cacao beans, producing a quasi-numerical expression, , which denoted a count of 40,000 cacao beans (Houston 1997). If this interpretation is correct and the bottom half of this glyph means 'unit of 8000 beans', then this is a technique for expressing large quantities that combines a bar and dot numeral with a sign for a metrological unit. If so, it provides a parallel to the interpretation of the calendrical period-glyphs as units equal to some quantity of a smaller unit (e.g., 1 tun = 360 days) but not read in terms of the smaller unit. In comparison with other civilizations, the Maya appear to have written large numbers very infrequently. That the Maya had these different means of writing larger numbers, and never used 'positional' vertical number columns in these non-calendrical contexts, casts doubt on the entire existence of positionality among the Maya.

The system of representing dates without period-glyphs is of great antiquity. As mentioned already, the technique was present in the epi-Olmec and Zapotec inscriptions by the 1st century BC, and continued to be used by the Preclassic Maya (Marcus 1976: 49-57). Although it was largely abandoned thereafter, Stela 1 at Pestac contains a date (9.11.12.9.0) written in this format, which refers to 665 AD (Closs 1986: 326-7). Most other Maya inscriptions include all the period-glyphs, although sometimes the glyph for the last position (kins) was omitted (Closs 1986: 308).

Our best evidence for the omission of period-glyphs comes not from stone monuments but from the Dresden Codex, a Postclassic text that was probably written in the early 13th century, though it may be a copy of a much earlier document (Marcus 1976: 35). It is the most astronomically sophisticated of the surviving Maya texts, and contains more of these vertical columns of numbers than any other. One of the strongest pieces of evidence suggesting the sophistication of Maya mathematics and astronomy are tables that have traditionally been defined as representing multiples of numbers such as 364, which Thompson believes were number of days in the 'computing year' used to calculate dates (Thompson 1941: 57). Nevertheless, the definition of the number 1.0.4 as 364, no matter how convenient it may be for Western scholars seeking to interpret Maya calculations, is conventional, and does not reflect specific knowledge of how the ancient Maya actually read such numbers. This is, alas, the only codex to contain such dates and constitutes the latest example of this sort of numerical notation, although there is one set of five numbers without period glyphs on the 15th century Madrid Codex that may qualify (Lounsbury 1978: 813). While the limited set of surviving texts in which this numeration was employed makes it difficult to trace its history, probably it was employed continuously throughout Maya history, largely on texts that have now unfortunately been lost forever.

In Chapter 8, I discussed the transformation of the Chinese traditional (multiplicative-additive) system into a ciphered-positional one by adding a zero-sign and deleting the exponent-signs for 10, 100, 1000, etc., so that 七千四十九 (7 x 1000 + 4 x 10 + 9) becomes 七〇四九 (7049). The astute reader will have seen already that there are similarities between this transformation and the removal of the Maya period-glyphs. The difference between the two is that in the Chinese case, the removed exponent-signs are *numerical* (representing the exponents of 10), whereas in the Maya case the period-glyphs are *calendrical*, not pure numbers. The assumption that the fourth position of the Maya numerals means "7,200" is wrong. It is particularly inappropriate to suggest, as various Mayanists have done, that numerals were written positionally in a purely vigesimal fashion for non-calendrical purposes - that is, with the third and fourth positions having the value of 400 and 8000 (Marcus 1976: 39; Lounsbury 1978: 764). The third (tun) position always denotes a period of 360 days, while the *katun* position represents units of 20 tuns, equivalent to 7,200 days.

Even so, it could be argued that if the katuns position does not mean "7,200", it could still have been read as "7,200 *days*". This is possible, but undemonstrated, and I do not consider it likely. When the period-glyphs are present, as they are in most of the inscriptions on stone, Mayanists do not consider the calendrical system to be a positional one, and do not treat dates as a sum of days. Why, then, should the removal of these period glyphs be anything more than an abbreviatory convenience? We recognize that one year is equal to 365 (or 366) days, but this does not mean that if I write the date "2002/06/14" I really mean a sum of days equal to 2002 years, 6 months, and 14 days, and certainly I do not calculate such a sum in my head. Granted, in the Maya calendar, where the Long Count is unaffected by leap years, months of different lengths, and other considerations, calculating a number of days is considerably easier than it is in the Gregorian calendar. Even so, why do we insist that the Maya must have been multiplying their dates out into sums of days?

The reason is that both historians of mathematics and Mayanists assume that positional notation was necessary, or at least highly useful, for doing calendrical calculations. Since the Maya obviously did do these calculations, and since these numbers look like positional notation, it is only natural to infer that they were read as such, despite the overwhelming evidence from the inscriptions on stone that dates were normally written as five different periods rather than as a single sum of days. Of course, the Maya undertook considerable feats of astronomy and calendrical computation. This suggests that at some point, they may have calculated using totals of days (the smallest calendrical unit with which they were concerned). Moreover, when Mayanists interpret Mayan chronology, they must translate Maya dates into a single number of days in order to the correlate Maya and Western calendars (e.g., the Goodman-Martinez-Thompson correlation establishes the beginning of the Maya Long Count as Julian day number 584,283). From this, it is easy to assume that since the Maya count can be interpreted etically as a count of days, the Mayan emic interpretation must also have been as a day-count.

Nevertheless, however the Maya may have read these columns of numbers, there is no evidence that they ever *calculated* with them. The Dresden Codex is a repository of calendrical data, including what appear to be multiplication tables, but there are no calculations on paper. In fact, there is specific ethnohistorical evidence concerning Maya computation, from Landa's *Relación de las cosas de Yucatan*, which suggests that the Maya did not calculate directly using bar and dot numerals:

> Their count is by fives up to twenty, and by twenties up to one hundred and by hundreds up to four hundred, and by four hundreds up to eight

> thousand; and they used this method of counting very often in the cacao trading. They have other very long counts and they extend them *in infinitum*, counting the number 8000 twenty times, which makes 160,000; then again this 160,000 by twenty, and so on multiplying by 20, until they reach a number which cannot be counted. They make their counts on the ground or on something smooth. (Tozzer 1941: 98)

I see no good reason to doubt Landa's assertion that computation was done on some sort of flat surface, which suggests that some sort of physical counting board was the primary computational technique employed, at least in the early 16th century. Some Mayanists have turned their attention to what sort of physical counters the Maya might have used and whether the bars and dots used as Maya numerals had physical correlates in rods and beans, or some other such markers (Tozzer 1941: 98; Thompson 1941: 42-43; Satterthwaite 1947: 30-31; Fulton 1979: 171). Sol Tax, working among the Maya of the Guatemala highlands at Panajachel in the 1930s, found that they used beans or stones in groups of five and twenty (though not with an abacus), supporting the idea that the ancient Maya may have done similarly (Thompson 1941: 42). Thompson's speculations on the functions of the Dresden multiplication tables, discussed above, are intrinsically tied to the theory that the Maya used a sort of abacus for calculating.

Counting-boards are quite often positional in structure, and some even use special counters or markers for empty positions - signs that resemble a zero in their function. Nevertheless, just as the Romans and Greeks had an abacus but no positional numerical notation system, we cannot assume that just because the Maya *may* have had a positional abacus, they *must* have also had positional numerals. The columns of an abacus work just as well if they indicate distinct units of baktuns, katuns, tuns, uinals, and kins as they do if they represent the exponent-values 144,000, 7,200, 360, 20, and 1. The manipulation of counters is identical, but the reading of the results is very different. While this is interesting speculation, it is impossible to confirm without further evidence, which we may hope will be forthcoming, given the vigorous pursuit of Maya studies at present.

Much less interesting are the host of speculations by scholars of numeration on the use of bar and dot numerals directly in calculation (Sanchez 1961; Bidwell 1967; Anderson 1971; Lambert *et al.* 1980; Mühlisch 1985). While, as Anderson (1971: 63) states, "it is not unreasonable to suggest that some attempt to use the numerals directly in computations might have occurred," this pastime tells us much more about the ingenuity of modern scholars than it does about the actual practices of Maya mathematics. Manipulating written numbers when computing, however common a practice in modern Western contexts, is a historical abnormality, which we have no evidence was relevant in the Maya case.

Unfortunately, the great bulk of Maya texts are now lost to us forever due to the tragic destruction of manuscripts on Spanish orders in the early colonial period. It is far too easy to create hypotheses concerning lost positional inscriptions when huge quantities of evidence have literally gone up in smoke. Yet the surviving evidence does not support the hypothesis that the number columns in the Dresden Codex should be interpreted as sums of days, and thus as a cumulative-positional numerical notation system. The most parsimonious explanation is that the omission of period-glyphs was abbreviatory but did *not* entail a radical re-reading of the numerical coefficients.

In his analysis of Maya arithmetic, Fulton noted that "it is possible to have a strictly positional notation, not altogether different from our present one, without any zero whatsoever" (1979: 171). In this, he is undoubtedly correct; positionality merely requires some way of avoiding ambiguity between, say, 749 and 7049, which may be simply an empty space, as it was in the Babylonian positional system prior to the Seleucid period (see ch. 7). Inverting this insight, I think it is probable that the Maya bar and dot numerical notation system has a zero, but does not use positional notation. This is not to say that the Maya zero or completion-sign was non-functional. While it was retained for aesthetic purposes in places where it was not strictly needed (when period-glyphs were present), zero-signs are needed whenever the period-glyphs are omitted and there is an 'empty' period. But the purpose of a Maya zero in a number such as 1.0.4 does not appear to be to indicate that the first number should be multiplied by 360, but rather simply to indicate that the middle position is empty, and thus the 1 should be read as 1 tun rather than 1 uinal. While something like positionality is used to distinguish different units of time, there was no Maya positional numerical notation system.

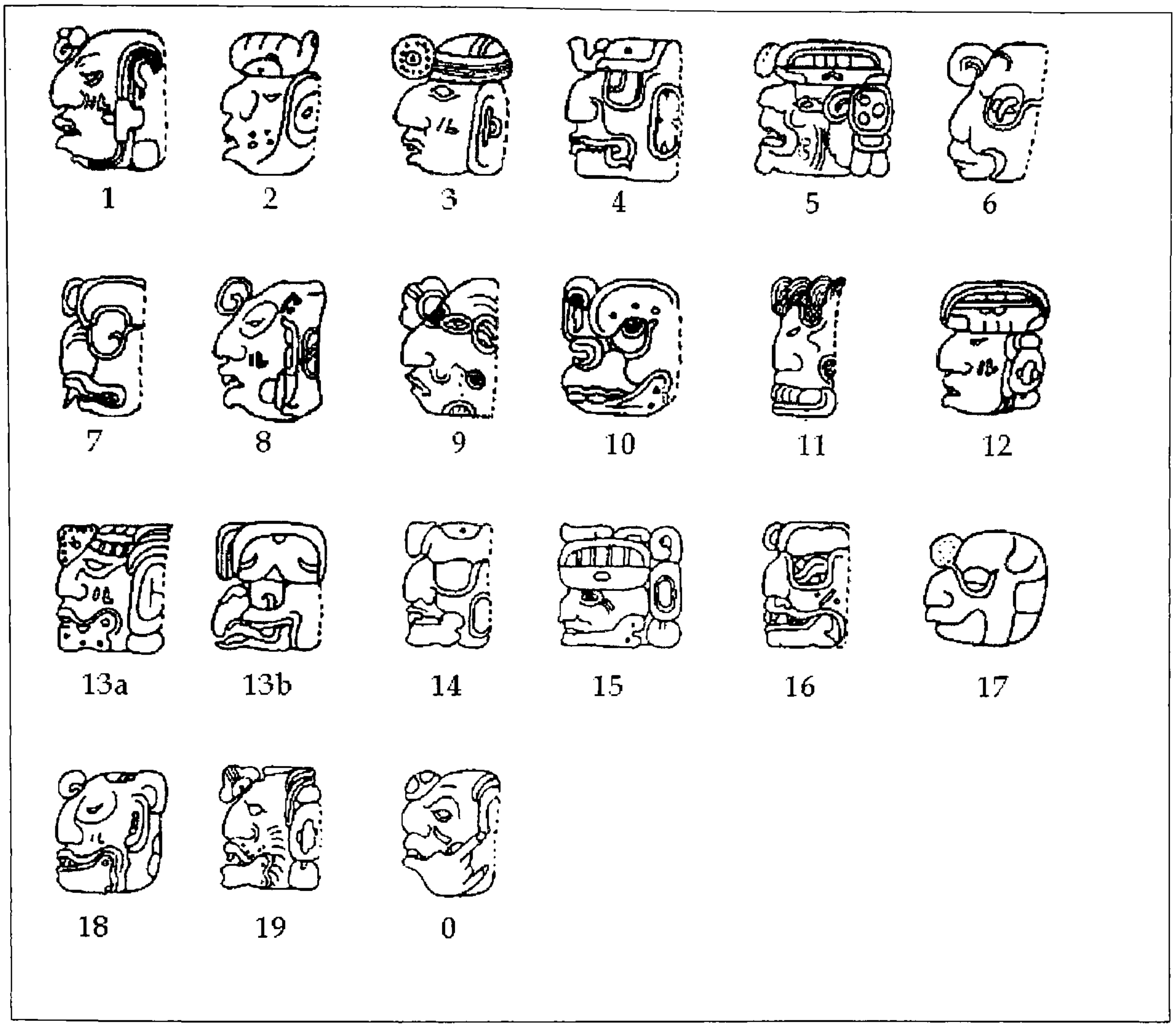
Maya head-variant numerals

In place of the bar and dot numerals, the Maya occasionally used a set of complex glyphs for the numbers 0 through 19, many of which correspond to the heads of Maya deities.[5] These head-variant glyphs are far more variable in form than are the very regular bar and dot numerals. Each head-variant could be used to replace the corresponding bar and dot numeral-phrase in an expression for a Maya date. An example of each of the signs is shown in Figure 9.5 (redrawn from Thompson 1971: Figure 24-25).

---

[5] For an analysis of the specific deities and other symbolism associated with each glyph, see Thompson 1971: 131-137; Macri 1982; Stross 1985.

Figure 9.5: Maya head-variant glyphs



Because the highest number that is expressed with head-variant numerals is 19, there is, strictly speaking, no vigesimal base to this system. However, in texts, head-variant glyphs are associated with one of the five calendrical coefficients (baktun, katun, tun, uinal, kin), and thus, when they occur in a single date, assume elements of a vigesimal structure. The head-variant numerals from 1 through 12 are written with elementary signs. The signs for 14 through 19 are additive combinations of the 'bared jawbone' element that represents 10 and the upper head of the sign for the appropriate unit. There are two signs for 13; the more common one (13a) is an additive combination of the bare jawbone for 10 and the head-glyph for 3, while the other (13b) is a distinct glyph for some sort of monster, and possibly holds some lunar significance as well (Macri

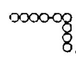
1982: 74). Because individual signs are not repeated to signify their addition, the head-variant numerals have more in common with ciphered than they do with cumulative numerical notation systems, but since they never exceed 19, they cannot be said to be either additive or positional.

Although the head-variants for 1-12 are elementary signs, it would be quite incorrect to state that this system has a base of 12; Kuttner's comments to the effect that the convenience of 12 as a base does not explain the 'duodecimal' head-variant system are thus quite misplaced (Kuttner 1986). The relevant sub-unit of the head-variant numerals is not 12 but 10, since the signs for 14 through 19 (and sometimes also 13) are expressed additively using 10 (Macri 1982: 75). No other Mesoamerican numerical notation system uses a decimal sub-base. The origin for this feature probably lies with the lexical numerals of the Maya languages, which use decimal structuring to express the numerals from 13-19, but have rather opaque formations for 11 (*buluc*) and 12 (*lahca*), just as the English "eleven" and "twelve" do not show any clear relation to "ten" (Lounsbury 1978: 762). Macri suggests that it may have been important to have 13 simple signs to correspond with the 13 deities used to name days in the Maya sacred calendar (Macri 1982: 48). It is conceivable that both of these explanations carry some weight, given the two variants for the 13-glyph.

The head-variant numerals are relatively common throughout the Maya stone inscriptions, though they are certainly less common than the bar and dot numerals. They also appear occasionally on the Dresden Codex, though not in any of the other codices (Thompson 1971: 131). They do not occur in any inscriptions from earlier than the Classic Maya period (ca. 250 AD) (Stross 1985: 37). Macri hypothesizes that they may have had a Preclassic origin, though the basis on which she does so is not clear (Macri 1982: 55). Regardless, unlike the bar and dot numerals, they were invented by the Maya themselves. Macri, pointing to phonetic correspondences between the head-variant signs and the lexical numerals of the Eastern Maya languages, suggests an Eastern Maya origin for the system (Macri 1982: 48). On the other hand, Stross, pointing out that many of the same correspondences exist with the Mixe-Zoquean languages spoken by the earlier Olmecs, suggests an Olmec origin for the head-variant glyphs (Stross 1985). This latter hypothesis seems unjustified, given that none of the epi-Olmec inscriptions contains head-variant numerals, and many centuries lie between the decline of the Olmec civilization and the appearance of head-variant glyphs. Given that the numerals are extremely different graphically and structurally from the bar and dot numerals, it would be fallacious to claim that the head-variants emerged from the bar and dot tradition. We had best think of them as a complex set of metaphors by which the numerical symbolism of deities was used as a sort of code for numerical information.

Given the destruction of so many Maya codices, as well as the imperfect state of Maya archaeology and hieroglyphic decipherment, it is difficult to say when the head-variant numerals ceased to be used. Since the Dresden Codex is the only codex to use them, and then only occasionally, it seems possible that they declined in use during the Postclassic period. I know of no post-Conquest evidence for their use.

Mexican dot-numerals

During the Maya Postclassic period (10th to 16th centuries), many of the peoples of central Mexico began using a system of dots to represent low integers in their pictographic manuscript tradition. Since this means of representation lacks a base and relies only on one-to-one correspondence, strictly speaking it does not constitute a numerical notation system. Nevertheless, it deserves some mention here, insofar as I think it is both descended from and ancestral to full-fledged systems. In their early history, the Mixtec and Teotihuacáni used bar and dot numerals, borrowed from the Maya or the Zapotecs, but after the 10th century AD, the system fell into disuse (Caso 1965: 955; Langley 1986: 143). While bar and dot numerals were occasionally used in a few later Mixtec codices, apparently for archaic or sacred reasons, they were largely replaced by a system whereby dots were used for the numbers 1 through 19, representing day-numbers and other objects (Colville 1985: 839-41). The peoples of Oaxaca, the Basin of Mexico, and the Gulf Coast used this system throughout the Postclassic period until the time of the Spanish conquest. A numeral-phrase was composed of a series of dots in a single row. To facilitate reading and to save space, larger numbers were often grouped in segments of three to five units, using connecting lines, changing direction of reading, or both, so that 9 might be written as . Numbers above 20 were never expressed in this pure-dot system.

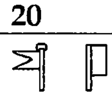
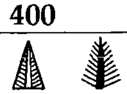
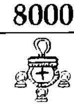
Given that the central Mexican calendar is part of a Mesoamerican calendrical tradition, and given the common use of dots for units in both the Maya and dot-only systems, I think it is plausible that between the 10th and 12th centuries AD, the use of bars for 5 was gradually abandoned, although the reason behind this change is not clear. The influence of Toltec culture, which was becoming predominant in Mesoamerica at this time, has been cited as the cause of this shift (Caso 1965: 955). Yet this argument begs the question of why the Toltecs did not adopt bar and dot numerals. Given that dot-only numerals are not known from anywhere in Mesoamerica prior to the 10th century AD, it seems unlikely that there was such a tradition prior to that point. Thus, unless we believe that the use of dots for units was independently invented in the two different parts of Mesoamerica, the dot-numerals must be descended from the bar and dot system.

It is also probable that the dot-numerals were ancestral to the Aztec numerals, which, as I will discuss below, constituted a base-20 cumulative-additive system. Because the Aztecs, like the Maya and Mixtecs, used dots for units, but because, unlike the bar and dot numerals, the Aztec system has no quinary component, the dot-numerals are a likely ancestor to the Aztec ones. Both the dot-numerals and the Aztec numerals use up to 19 dots for units, the difference being that with the Aztec numerals, the dots were more regularly grouped in fives, and higher numbers were written using different signs for the exponents of 20. It is generally believed that the Aztecs inherited their tradition of manuscript writing from the Mixtecs (Colville 1985: 839). Dot-numerals continued to be used in Aztec manuscripts even after the development of the cumulative-additive numerals in the 14th century. It therefore seems very improbable that the Aztec numerals could have developed entirely independently. Because the chronology of the development of these systems remains unclear, the exact relationship between them remains hypothetical. By the time of the Spanish conquest, the Aztec numerals had supplanted the dot-numerals in some areas outside their tributary area, and were used in many of the post-Conquest Mixtec codices (Terraciano 2001).

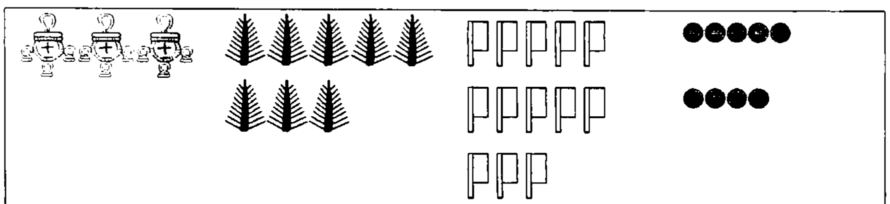
Aztec

The name 'Aztec' applies most precisely to the Nahua-speaking inhabitants of the region immediately surrounding the ancient city of Tenochtitlan (modern Mexico City), who controlled a substantial tributary system in central Mexico between the 14th and 16th centuries. More generally, the term is often used to refer to the various Uto-Aztecan-speaking peoples of central Mexico who were under Nahua rule during this period. The Aztec tributary network, which embraced numerous small states, produced a large number of manuscripts, using a combination of ideographic and phonographic signs. The considerable debate concerning whether this Aztec manuscript tradition constituted true writing or simply served as a semasiographic mnemonic aid is irrelevant to the study of Aztec numeration. The Aztecs most definitely possessed a vigesimal numerical notation system, whose signs are shown in Table 9.2.

Table 9.2: Aztec numerals

1	20	400	8000
●			

The sign for 1 is obviously the dot that was commonly used for units throughout Mesoamerica. The signs for the vigesimal exponents are depictions of objects: for 20, a flag (*pantli*), for 400, a feather (*tzontli*, literally 'hairs'), and for 8000, a bag used to hold copal incense (*xiquipilli*) (Harvey 1982: 190). These signs were combined in a cumulative-additive fashion, normally written in horizontal rows with the highest exponents on the left. While the Aztec numerals, unlike the Maya bar and dot numerals, did not use a sign for 5, groups of more than five identical signs were arranged in groups of five for easier reading. In addition, groups of five signs were sometimes joined to one another with a horizontal line underneath the set. Thus, 27469 might be written as shown in Figure 9.6.



(3 x 8000) + (8 x 400) + (13x20) + (9 x 1)

Figure 9.6: Aztec numeral-phrase for 27469

The purely vigesimal structure of the Aztec numerical notation system and the shapes of its numeral-signs are quite different from the lowland Mesoamerican bar and dot system. As I have just argued, the Mexican dot numerals are the most likely ancestor of the Aztec system. It seems quite plausible that the Aztecs originally used dots alone, but then, as the administrative needs of their tributary system grew, invented new numeral-signs for 20 and its exponents. As far as can be discerned, however, the inventors and early users of the Aztec system were unaware or uninfluenced by the lowland Mesoamerican systems of the Maya. Because the Mexican dot numerals do not constitute a numerical notation system according to my definition, it is important to recognize that the Aztec system was invented relatively independently of influence from other numerical notation systems.

One of the most important functions of the Aztec numerals was to record the results of economic transactions, such as amounts of cacao beans, grain, clothing, and other goods received from different regions of their tributary system (Payne and Closs 1986: 226-230). Numerals were also used in Aztec annals and historical documents, such as the record of the massacre of 20,000 prisoners in the Codex Telleriano-Remensis (Boone 2000: 43). Sometimes, when recording amounts of goods, individual numeral-signs were attached to an equal number of pictographic signs for goods. Accordingly, one might record 1200 balls of incense not as the numeral 1200 () followed by a picture of an incense ball, but rather using three balls of incense, each of which would be placed immediately underneath a sign for 400.

The use of Aztec numerals to record large quantities of tribute and individuals stands in sharp contrast to the Maya bar and dot numerals, which were almost wholly calendrical in function. The Aztecs denoted their 13 months using series of dots in rows, just as the Mixtecs did, but when they did so, they did not group dots regularly in groups of five (Boone 2000: 43-44). These should thus be considered as a continuation of the dot-numerals in Aztec manuscripts rather than being part of the Aztec numerical notation system. Normally, the Aztecs did not record dates or other calendrical information using the larger numeral-signs. In a single text, the Vatican Codex, large periods of time seem to have been expressed using cumulative-additive combinations of different signs, the largest of which represents 5206 years with 13 signs that probably represent 400 () above which six dots were written (Payne and Closs 1986: 234-5).

After the Spanish conquest, the Aztec numerical notation system continued to be used in various colonial documents. In fact, its use spread beyond the regions traditionally under Aztec control, as Nahuatl increasingly became a *lingua franca* used by the non-Spanish inhabitants of the region. For instance, Aztec numerals are common in the Mixtec Codex Sierra, a mid-16th century account book that uses Western, Roman, and

Aztec numerals side-by-side (Terraciano 2001: 40-45).[6] In a few post-conquest manuscripts, fractions could be depicted by segments of 1/4, 1/2 and 3/4 of a dot, low multiples of five by filling in quarters of the *pantli* flag sign, and 100, 200 and 300 could be expressed using segments of the *tzontli* sign for 400: , , and  (Vaillant 1950: 202).

An interesting development in the Aztec numerical notation system is the use of multiplicative rather than strictly additive notation in a few post-Conquest codices. Guitel (1958; 1975: 177) was the first to point out that one of the often-reprinted examples of Aztec numbers depicts a basket of cacao beans from which 4 signs for 400 emerge, above which a *pantli* or flag for 20 is placed. This numeral-phrase represents a total amount of 32,000 cacao beans multiplicatively, as 20 baskets of 1,600 beans each, rather than additively, as 4 *xiquipilli* of 8,000. I am unconvinced of the significance of this single numeral-phrase, because in a circumstance where cacao beans come in baskets of 1,600 beans, it seems important to denote that there are 20 baskets of 1,600 each, not simply "32,000 beans". This does not certify that placing the numeral-phrases for 20 and 1,600 together means "32,000". However, Guitel was not aware of another text, a Texcocan document now known as the Codex Kingsborough, where multiplicative notation was used extensively (Paso y Troncoso 1912; Harvey 1982). As this system is a structurally distinct variant of the standard Aztec system, I will treat it separately below.

As disease, warfare, and acculturation diminished the strength of Aztec traditions, the old numerals ceased to be used. I do not know of any documents from later than 1600 that use Aztec numerals. After this point, Roman and especially Western numerals were employed throughout highland Mexico.

[6] Boone (2000: 254) indicates that Oaxacan texts do not contain signs for 400 or 8000; at least in the case of the *tzontli* sign for 400, she is incorrect, as this is found in the Codex Sierra (cf. Terraciano 2001: Figure 2.16).

Texcocan line and dot

The city of Texcoco in the province of Tepetlaoztoc was one of the most powerful cities in the Basin of Mexico both before and after the Spanish conquest of the Aztecs. While many colonial documents of the 16th century continued to use the Aztec numerals described above, a handful of Texcocan documents contain a rather unusual numerical notation system. This system, which I will call the 'Texcocan line and dot' system, has been studied extensively by Harvey and Williams (Harvey and Williams 1980, 1981, 1986; Harvey 1982; Williams and Harvey 1997). The numeral-signs of this system are shown in Table 9.3 (Williams and Harvey 1980: 500).

Table 9.3: Texcocan line and dot numerals

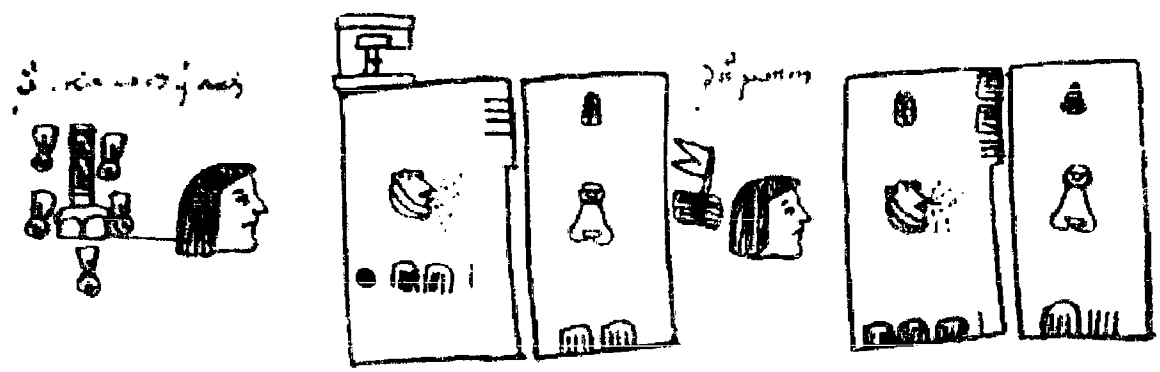
1	5	20
|		●

This system is cumulative-additive, with a base of 20 and a sub-base of 5. The 'sign' for 5 is no more than five unit-strokes joined together by a curved line, so it is perhaps just a matter of personal preference whether we see it as a separate numeral-sign. A similar technique sometimes was used to group sets of five dots for 20 as , but this was not universally done. Perhaps the most unusual feature of this system is that, whereas other Mesoamerican numerical notation systems used a dot for the units, here a vertical stroke was used for the units, while the dot took a value of 20. Numeral-phrases could be written in any direction (vertically, horizontally, forwards, and backwards), but were always arranged in a single line from highest to lowest signs (Harvey 1982: 191). Thus, 72 could be written as  or , but not .

This form of notation has been found on only three texts, all of which were written in the vicinity of Texcoco in the 1540s. Two of these, the Códice Vergara and the Códice de Santa Maria Asunción, were cadastral records written around 1545 to enumerate individuals and their land-holdings. These two are in fact so similar that they may have been parts of the same manuscript at one point, or at least were drawn at the same time (Williams and Harvey 1997: 2). The third, the Oztoticpac Lands Map, was written around 1540, and is also a record of lands, though as a map rather than a census record (Cline 1966). The primary function of the numerals in all of these cases was to record land measures, which raises the possibility that the line-and-dot system may not have recorded abstract numbers but rather linear units of measurement. Indeed, pictographic signs expressing fractional linear units sometimes accompanied the numerals, but their meanings are still unclear (Williams and Harvey 1997: 26). Nevertheless, several numeral-phrases on the Oztoticpac map were used to count sums of days, so we can clearly state that this was a true numerical notation system.

The Códice de Santa Maria Asuncion is especially interesting in that it appears to use a modified form of this system to express numbers positionally rather than additively. In studying this text, Harvey and Williams (1980) showed that line and dot numerals occurred in two different sections, but served very different functions. In one section, known as *milcocoli*, line and dot numerals were used in the regular manner by writing them along the edges of maps of plots of land owned by different individuals. In the milcocoli section, numerals indicated the lengths of various sides of plots of land. In another section, known as *tlahuelmatli*, the plots of land from the milcocoli section were redrawn as rectangles (regardless of their original shape). While this section also contained line-and-dot numerals, they did not indicate linear units but rather the areal measurement of each individual's land holdings. By comparing the milcocoli values, which indicated the lengths of the sides of plots, and the tlahuelmatli values, which recorded their total area, Harvey and Williams discovered that a form of positional notation was used to record land areas in the tlahuelmatli section of the codex using a set of three registers in which numbers were written. A sample of this notation is shown in Figure 9.7 (Harvey and Williams 1986: 242).

Figure 9.7: Positional numerals in the Códice de Santa Maria Asunción

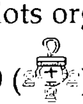


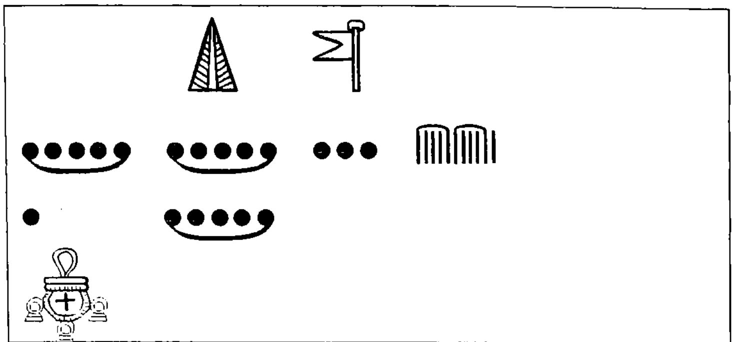
31 x 20 + 4 = 624 | 10 x 20 = 200 | 16 x 20 + 13 = 333 | 9 x 20 = 180

In the top right corner, values from 1 to 19 were indicated using line and dot numerals in a small protuberance (seen in the first and third plots of Figure 9.7, where 4 and 13 are denoted). On the bottom line of the rectangle, units and groups of five indicated multiples of 20 units (10 x 20, 16 x 20, and 9 x 20, respectively, in the last three plots). No dots were ever used in either of these two registers. When dots were found, they occurred with or without units in the centre of the rectangle. Strangely, this third register also counted multiples of 20 (i.e. lines equal 20 and dots equal 400). No plots of land show values both on the bottom line and in the centre of the rectangle. When the twenties register and the units register were added together, a total area value was reached; the four plots in Figure 6 denote 624, 200, 333, and 180, respectively. Harvey and Williams found that in 71% of the land plots they examined, the tlahuelmatli value was within 10% of the projected area for that plot based on the milcocoli measures (1980: 501). While this may not seem to be remarkably accurate, we must remember that the plots were often very erratic in shape and that calculating area was not simply a matter of multiplying length by width. Harvey and Williams (1980: 501) also show that where there is no value in the third (central) register, a corn glyph, or *cintli*, is drawn at the top of the rectangle. This sign, which can be seen in the second, third, and fourth plots in

Figure 9.7, may indicate that the third register is empty, and thus serves one of the functions of a zero-sign.

Taken as a whole, these numbers can be read as a base-20 cumulative-positional numerical notation system with a sub-base of 5. Unlike Western numerals, in which the positions are arranged in a straight horizontal line, the Texcocan system uses three registers, the last two of which have an identical positional multiplier. I am unconvinced that the *cintli* glyph really serves as a zero, because it does not indicate an empty exponential position, but rather provides information as to where to find the twenties exponent (on the bottom line, rather than in the centre of the rectangle). While I think that the correlation established by Harvey and Williams demonstrates that the tlahuelmatli value represents an area value, I am not fully convinced that it is meant to be read as a single number, but may instead represent two values, one of which represents a larger area value that is twenty times another value. I do not know how this issue could be resolved at present.

A unique Texcocan document from 1555, the Codex Kingsborough, also uses something like the line and dot numerals (Paso y Troncoso 1912). It was a record prepared as part of a lawsuit, and denoted the amount of tribute paid to Spanish officials by the inhabitants of the Tepetlaoztoc region, with extensive description in Spanish that confirms the numerical values (Harvey 1982: 193). The most interesting feature of this text is its use of a combination of the Aztec pictographic numerals and the Texcocan line and dot notation. Lines and chunked groups of five lines indicate 1 and 5, respectively. To write higher numbers, dots organized in lines of five were placed beside the signs for 20 (), 400 (), and 8,000 (). The dots were placed in a single row, with the signs for 20 and 400 above them and the 8,000 sign below them. Thus, where the regular Aztec numerals use these three signs cumulatively, the Kingsborough numerals are written using just one of each sign, next to which units from 1 to 19 were expressed with dots. Thus, 52,071 might be written as shown in Figure 9.8.



		400		20
6 x	+	10 x	+	3 x + 5 + 5 + 1
8000				

Figure 9.8: Codex Kingsborough numeral-phrase for 52,071

Whereas the basic line and dot system is cumulative-additive, and the tlahuelmatli system is cumulative-positional, this system is multiplicative-additive. The total value of the numeral-phrase is taken by multiplying the dots for units by the values of the exponent signs and taking the sum. However, while the dots look like the "20" dots of the line and dot system, they each stand for 1 in this system. To add to the complexity of this situation, in some cases the flag glyph for 20 could be omitted, retaining only the dots (Paso y Troncoso 1912: 261r, 238v, etc.). In these cases, we have the elements of a cumulative-positional system, since the value of the twenties exponent is determined by its position in the numeral-phrase through implied multiplication. Finally, in a couple of numeral-phrases, lines are placed to the left of dots, as where a number is written as ||●●, which might be read as 42 (from right to left), but could have a variety of other interpretations (Paso y Troncoso 1912: 274v). The erratic nature of the numerals used in this text suggests that whoever wrote it was extremely inventive and was in the process of experimenting with different means of representation.

The most important question regarding the line and dot numerals, their positional variant in the tlahuelmatli records, and their multiplicative variants in the Codex Kingsborough, is whether these systems existed before the Conquest, or if their development was stimulated by contact with the Spanish. Neither the Western or Roman numerals are cumulative-positional or multiplicative-additive, and neither uses a base of 20, so the Texcocan systems are structurally distinct from those of the Europeans. Thus, it would be premature to conclude that Spanish contact brought about the development of these systems. Harvey and Williams (1980: 503) argue that, while the tlahuelmatli numerals are positional and have something like a zero, the use of different registers around a rectangle is quite different from Western positionality, and the zero does not serve the same functions as the Western zero. On this basis, they regard it as a native invention.

I believe they are essentially correct, and that the line and dot numerals and their variants are so different from Western and Roman numerals that they could not have been introduced by the Spanish. Nevertheless, these may be instances of stimulus diffusion, which the Texcocan scribes developed with an awareness of Western and/or Roman numerals but without adopting the form and structure of those systems. That the Texcocan systems occur only in a handful of documents in a single region in the generation immediately after the Conquest and cease to be used after only two decades suggests that this was not a system of great antiquity. At the very least, it seems likely that the multiplicative (Kingsborough) and positional (tlahuelmatli) variants were stimulated by contact with the West, while the cumulative-additive line and dot numerals may well have existed prior to Spanish colonial rule. Because the study of the Mexican documentary evidence from the 15th and 16th centuries is still quite spotty, there is every hope that further research might resolve the issue.

Other systems

Because our understanding of Mesoamerican numerals is imperfect, a number of Mesoamericanists have developed theories regarding other forms of written numeration whose existence I cannot confirm. I think it quite likely that more numerical information has been recorded than we are presently able to read in the Maya, Zapotec, Teotihuacáni, and Aztec texts. Even if the hypotheses below turn out to be incorrect, some elements of them may be salvaged in the reconstruction of as-yet unknown numerical notation systems.

Urcid Serrano reports on an extremely peculiar hypothesis of Howard Leigh, who in the 1950s postulated that the Zapotec inscriptions contained encoded astronomical data, and thereby assigned numerical values to various glyphs (Urcid Serrano 2001: 49-50, 54). In addition to bars and dots, this supposed system had over 20 unique signs. These include elements of base-10, base-13, and base-20 notation, culminating in a special sign for 1,186,380 (3x3x3x13x13x13x20), and signs for the number of days in the synodical revolutions of the planets Venus (585) and Mars (780). I am unconvinced that such a system actually existed. I mention it because the Zapotecs may have encoded numerical information in some of these glyphs, though not in the way Leigh imagined them to have done.

It is also possible that an unusual cumulative-additive bar and dot numerical notation system existed at Teotihuacán, one in which the bars did not have a fixed value but could mean 5, 10, or 30, depending on their configuration (Langley 1986: 141). The nature of the script of Teotihuacán is still controversial, though it is increasingly thought that there was a complex pictographic script of the type used later in highland Mexico (Taube 2000). However, because Teotihuacán never used phonetic writing, and because, unlike the Aztec situation, there is no body of colonial documents to explain the numerals, there is no way to confirm the values of any potential numeral-signs.

Penrose (1984) asserts that in the almanac portions of the Dresden, Madrid, and Paris codices, the Maya used "cryptoquantum" numerations to represent an encoded quantity of days in a manner quite distinct from the bar and dot or head-variant numerals. He argues that the Maya represented hidden counts of large numbers, by assigning numerical values to special signs indicating the days of the "Sacred Round"

260-day calendar, and then by manipulating them through multiplication. Mayanists do not appear to be aware of Penrose's research and his conclusions must be viewed as highly speculative at present. I am unconvinced that the manipulations necessary to extract meaningful numerical information from these signs are anything more than numerological play.

An unusual form of numerical notation is employed on the Codex Mariano Jimenez, a 16th century post-Conquest manuscript from Otlazpan (in the province of Atotonilco). It uses dots for units, horizontal lines for twenties, and horizontally oriented *tzontli* (feather) glyphs for 400, with fractions of 400 depicted by showing partially denuded feather-signs. The system seems to be purely cumulative-additive. Although treated by Harvey and Williams (1986: 251-253) as simply a variation on the Texcocan system described above, the differences between the two systems suggest that they are quite distinct. If more documents using this sort of notation are found, it may be possible to confirm the existence of yet another post-Conquest regional variation of the Aztec numerals.

Summary

The two features common to all the Mesoamerican numerical notation systems is that they have a vigesimal base and that they are all cumulative rather than ciphered. The Maya head-variant glyphs, a sort of ciphered system that only expresses units, constitute a partial exception to this rule, but, being a sort of symbolic code, they are inherently quite unusual numerals. The presence of a quinary element is quite common, as is the use of dots for units, but neither of these features is found throughout the family. Like the East Asian family (ch. 8), Mesoamerican numerical notation systems use a variety of basic principles, and our primary evidence for their unity as a family is historical rather than structural.

The assumption that the numerals of the region are fully understood needs to be challenged persistently. When the bar and dot numerals were the only part of the Maya script to be deciphered, it must have seemed remarkable indeed to be able to extract calendrical information from such otherwise inscrutable documents. Yet our understanding of the cultural history of the Mesoamerican numerical notation systems is less complete than for most Old World families. As our reading of Maya and Aztec writings becomes more sophisticated, it is to be hoped that we will come to a clearer understanding of their numerical notation.

Chapter 10: Miscellaneous Systems

In this chapter, I describe around twenty systems that do not fit neatly into any of the phylogenies of chapters 2 through 9. Nevertheless, these systems can be grouped into several categories on the basis of their origins. A few, such as the Inka *quipu*, the Indus script numerals, and the enigmatic Bambara system, arose independently of any other system and apparently gave rise to no descendants. Others are cryptographic or limited-purpose systems used in medieval and early modern manuscripts of Europe and the Middle East. The majority, however, developed in the past century, and their origins are well understood. These systems emerged in colonial settings, usually under the influence of the Western or Arabic ciphered-positional numerals and in conjunction with the development of an indigenous script. Most of these systems were developed in sub-Saharan Africa, but Asian (Pahawh Hmong, Varaŋg Kshiti) and North American (Cherokee, Iñupiaq) indigenous groups have also developed their own numerical notation systems. Because none of these systems is clearly related to another, I have not included a comparative chart as I have in previous chapters.

Inka

The Inka civilization, which controlled an enormous state on the west coast of South America between 1438 and 1532, lacked a writing system capable of expressing phonetic values. Instead, the primary means of encoding[1] information was a system of knotted ropes of different colours, known as *quipus*, which recorded numerical quantities. About 500 to 600 quipus survive, although accurate provenances cannot be established for many of them (Urton 1998: 410). Despite their obvious numerical function, which was established by Locke (1912) nearly a century ago, quipus are often considered to be

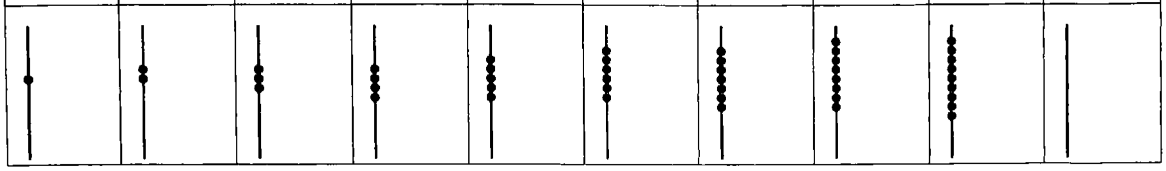
[1] I use the term "encode" instead of "write" when discussing the quipu notation; however, this does not imply that I consider the quipu to be fundamentally different from other numerical notation systems that are 'written'.

qualitatively different from written numerals, and are lumped together with unstructured systems that use one knot for one object, a practice that is widely attested ethnographically (cf. Ifrah 1998: 70). As Ascher and Ascher (1972: 292) note, the confusion between simple tallying objects that use one-to-one correspondence and complex systems for representing higher numbers plagues the study of quipus by non-Andeanists. Quipus contain a numerical notation system (a positional one, in fact) and should be compared to written numerals rather than to simple tallies using knots. The fact that they are not written is largely irrelevant.

A quipu is a set of coloured cotton or wool cords consisting of a main cord (up to a metre in length) from which multiple cords containing knot-numerals are suspended. The numeral-bearing cords are subdivided into pendant cords, which hang directly down from the main cord when it is held horizontally and stretched taut, top cords, which hang from the main cord but are tied so as to lay on the opposite side of the pendant cords, and subsidiary cords, which hang from a pendant cord, top cord, or another subsidiary cord rather than the main cord (Ascher and Ascher 1980: 15-17). The designation that pendant cords hang "down" and top cords hang "up" is an artifice; while they naturally hang on opposite sides of the main cord, we do not know how they would have faced when used by the Inka. Each of these cords usually contains a numeral-phrase, or, more rarely, two. The system is cumulative-positional with a base of 10. In each position, the value of that exponent of 10 is encoded using between zero and nine knots or loops. The units position is the one farthest from the main cord (its loose end), while the highest exponent is found closest to the main cord. While a quipu theoretically could express any number (because the system is positional), in practice, five-digit numbers are the highest recorded, and these are quite rare (Ascher and Ascher 1972: 291). When a position is empty, there is no sign for zero; instead, an empty space was left on the cord to permit its correct interpretation. Table 10.1 shows the quipu-signs, depicted in a stylized form.

Table 10.1: Inka quipu numerals

1	2	3	4	5	6	7	8	9	0



It is slightly misleading to depict these representations as sets of identical dots, because three distinct knots were used, depending on the number being represented; these are depicted in Figure 10.1 below (Ascher and Ascher 1980: 29). To encode a value in the tens, hundreds, or higher exponents, the quipu maker (*quipucamayoc*) would tie an appropriate number of single knots in a line. For the ones exponent, however, two different types of knot were used. For all the units except 1, a long knot was used in which the cord was looped around itself an appropriate number of times for the number being expressed; the knot shown in Figure 1 thus represents 4. Because a long knot cannot be made with fewer than two loops, a value of one in the units position required the use of a different knot, a figure-8. The use of different knots might appear to take away from the purely positional nature of the system. Yet, because there is no zero-sign, this technique greatly reduced the chance of misreading a cord. If a cord contained 6 single knots followed by 2 single knots, it could not be read as 62 but only as 620 (or possibly 6200). The use of long or figure-8 knots in the units position makes it much easier to tell which is the units position, and thus to identify the subsequent positions.

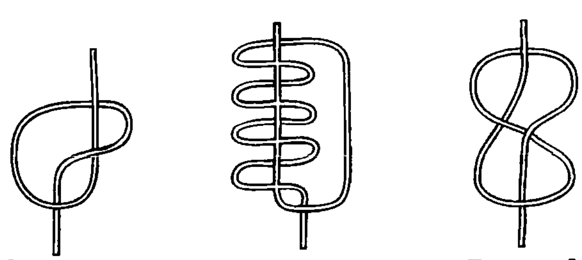


Figure 10.1: Quipu knots

Figure 10.2 depicts an example of an unattested but plausible quipu that reflects the notation of numbers on various cords. The main cord lies horizontally, with the pendant cords (P1 through P4) hanging down and the top cord (T1) facing up, with subsidiary cords (S1 through S3) hanging off both pendant and top cords. On this cord, only a single value would have a figure-8 knot (the 1 in the units position on P4); the other units values (3 on P2, 6 on S1, 2 on P3, 6 on T1, and 6 on S3) would be made with long knots, and all the tens and hundreds figures with single knots. As is sometimes the case in attested quipus, the top cord value (776) is equal to the sum of the pendant cords (360+23+102+291), while the value on the top cord's subsidiary (S3 = 26) is the sum of the subsidiaries of the pendant cords (20+6).

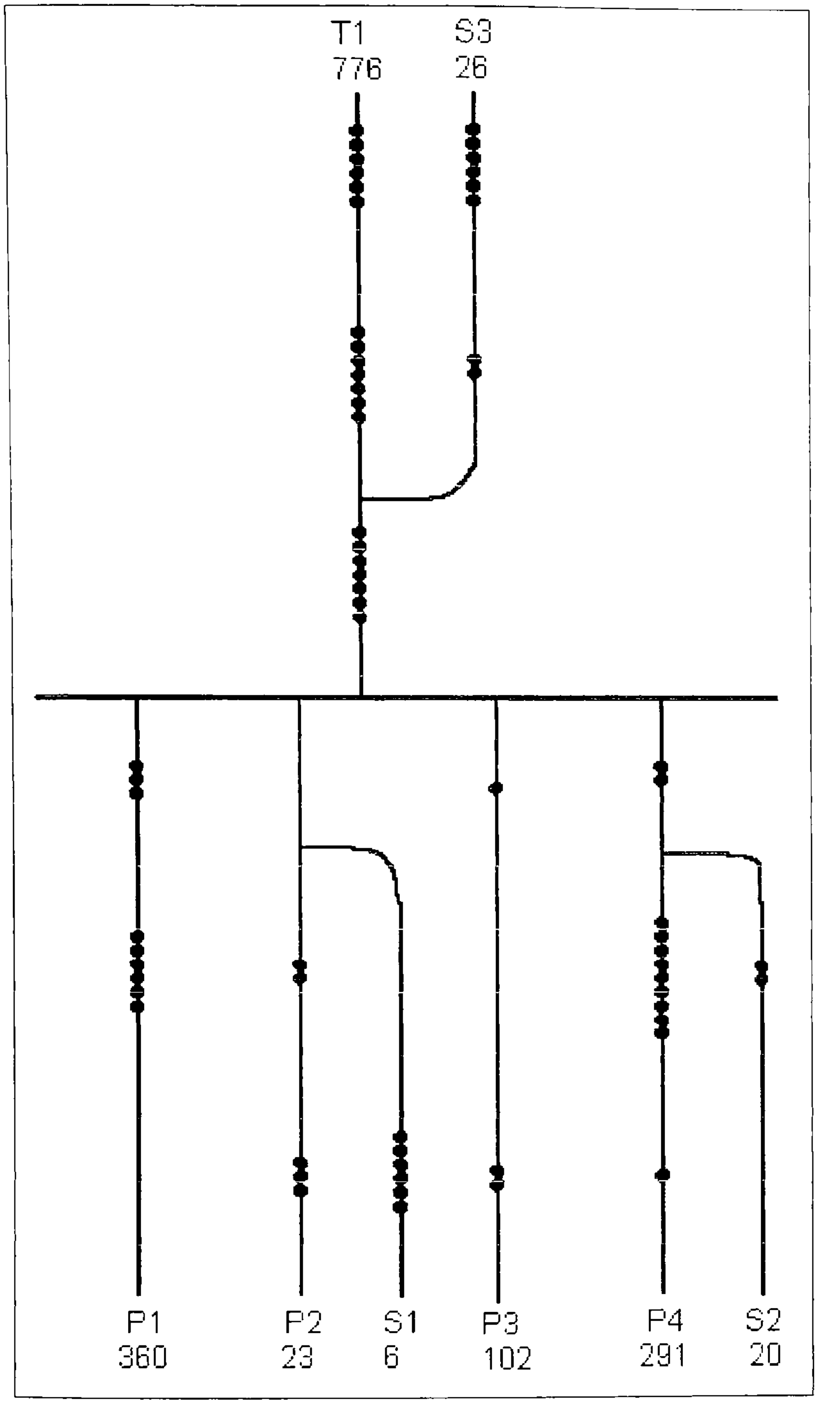


Figure 10.2: Quipu structure

Although we can read the numerical values on quipus, we know very little about their origin and early history. We can assert with confidence that they were used throughout the period of Inka dominance in the Andes (starting in the early 15th century), but we do not know how much earlier they were used. Most of the quipus surviving today were collected haphazardly; there are only two archaeological discoveries of quipus with adequate provenience (Urton 2001: 131). Bennett (1963: 616) notes that some early Mochica vessels bear markings that are suggestive of quipus. Yet we have no direct evidence for their use in the Chimú empire or any of the other pre-Inka Andean states, so it is probably best to see their invention as a strictly Inka phenomenon. It is possible that the quipu system developed out of an earlier knot-based system using simple one-to-one correspondence, but there is no evidence for this save that knot tallies of this sort have a very wide distribution in the Circum-Pacific region (Birket-Smith 1966).

The functions for which the quipus were used are fairly well known. Their primary use was as the record-keeping system of the Inka state; they were employed in this capacity for censuses, tributary records, and similar administrative functions. The decimal base of the quipu notation system corresponds with the decimal divisions of society by which the state was administered. It is also known that some quipus contained calendrical information (Ascher and Ascher 1989, Urton 2001). For instance, quipu UR6 from the Laguna de los Cóndores site contains a series of cords with values of 20 to 22 followed by cords with values of 8 or 9, and the sum total of these cords is 730 (365 x 2), strongly suggesting that it may have been a biennial calendar (Urton 2001: 138-143). It is suggested from ethnohistorical data that genealogical, historical, and literary information might also have been recorded using quipus (Bennett 1963: 618). While multiple early chroniclers reported such a function for the quipus, how this would have been done is unclear. It is also important to note that most surviving quipus have been recovered from grave sites or tombs. It is probable that the Inka placed quipus in the graves of *quipucamayocs*. It is unclear whether this implies that some of them should be read as 'tomb texts', because presently we are unable to extract non-numerical information from them (Urton 2001: 34).

Much ink has been spilled recently about whether the use of knots on quipus constituted something more than a numerical notation system, and whether we should regard it as approximating the functions of a writing system. Gary Urton (1997, 1998,

2001) has argued forcefully that many quipus contain syntactic and semantic information possibly far exceeding their numerical functions. While not denying that they can be read as representing decimal numerals, he contends that purely numerical readings that translate quipu texts as Western numerals "inevitably mask, and eliminate from analysis, any values and meanings that may have been attached to these numbers by the Quechua-speaking bureaucrats of the Inka empire who recorded the information" (Urton 1997: 2). He argues, quite rightly, against the idea that a quipu could only have been interpreted by its maker or those trained in an idiosyncratic private code (Urton 1998: 412). It is also obvious that the quipus must have recorded more than simply numerical information; a list of pure numbers is practically useless. In some way, at least the nature of what was being counted must have been recorded somehow. The most likely possibility is that this was done with colour; the post-colonial chronicles of Garcilaso de la Vega, one of the more reliable of the Hispanic-Inka chroniclers, inform us that coloured cords were used to record different quantities (Bennett 1963: 617). Unfortunately for this theory, many quipus use multiple colours of cord, and no reliable means of reading the type of items counted has yet been developed.

It would be folly, at this stage in the interpretation of the quipus, to cut off any avenue of investigation entirely. I think a case can be made that the quipus encode approximately the same amount of information as the proto-cuneiform accounting signs of Mesopotamia (ch. 7), which identify only items being counted and the quantity of each item. Since the proto-cuneiform system is regarded as 'proto-writing' or even by some authors as the world's first 'script', I think it is reasonable to attribute the same status to the Inka recording system, especially since there is abundant ethnohistorical evidence to support this assertion (Urton 1998: 417). Furthermore, I think it possible that the quipu system, over time, might have developed into a system for representing speech (though it is probably more difficult for a knot-based notation than it is for a system based on inked or impressed signs).

Nevertheless, I am unconvinced that the quipus ever encoded a writing system capable of expressing narratives. Unless we believe that some quipus are entirely non-numerical (and I know of no one who has contended this), the numerical presence in the quipus is inescapable. I cannot see how a single quipu can be at the same time both a record of numbers and of things being enumerated *and* a fully developed system for recording history and literature. Urton's (1997: 179) speculation that there might have been two pre-colonial quipu systems (one for recording quantity and another for recording narrative) lacks an empirical foundation. He is well aware that all the post-Conquest chroniclers state explicitly that the quipus carried only numerical meanings. Therefore, Urton postulates that the early colonial Spanish, in order to undermine traditional patterns of knowledge, rapidly transformed the quipu system from a full-fledged writing system into a purely numerical and non-narrative recording instrument (Urton 1998: 410-411). I admit that the Spanish may have wished to denigrate Inka knowledge, and also that there is an enormous issue of translation between indigenous concepts and what is claimed in early Spanish chronicles. Yet there is no direct evidence to support Urton's proposition. It would have been much simpler to replace the quipu system with European administrative techniques than to attempt such an alteration of its function. Finally, I am unconvinced that analogies with the mathematical practices of modern Quechua-speaking peoples, such as those drawn by Urton (1997), will help us further to interpret centuries-old quipus unless continuity can be demonstrated (not merely assumed) between pre-colonial and modern ways of thinking. Barring the discovery of non-numerical quipus or other Andean recording systems, I think it is far simpler to interpret the quipu system as a "number + noun" information system, of which only the numerical component can be determined in most cases. The earliest Mesopotamian civilization did not require phonetic writing, nor did that of the Yoruba, to mention only two highly complex but non-literate sets of polities. To infer a writing system out of nothing but an assumption that such a system would have been necessary is grossly anti-empirical.[2]

Whatever the quipus may or may not have been, it is evident that they alone cannot have been used for performing arithmetical calculations. Quipus are even less amenable to physical manipulation than are written numerals (which can be lined up and crossed out). We do know, from 16th century documents, that the *quipucamayoc* were responsible not only for making and reading the quipus but also for calculating the results that they would then encode on the cords, and that they did so using a set of stone tokens (Urton 1998; Fossa 2000). While no archaeological evidence has confirmed the existence of such a system, there is limited documentary evidence for an "Inka abacus" in a document written between 1583 and 1613 by Don Felipe Guaman Poma de Ayala (ca. 1534 - 1615), a descendant of an Inka princess who was an important chronicler of life in late 16th century Peru and a critic of Spanish rule (Wassén 1931; Urton 1997: 201-208). In one corner of a page depicting a *quipucamayoc* at work, there is a grid of five rows by four columns, in each square of which is found a number of circles: five dots in the first column, three in the second, two in the third, and a single dot in the fourth. Moreover, some of the dots have been filled in while others remain empty. Unfortunately, while the commentary that accompanies this picture clearly notes that the Inka reckoners used computing boards, there is no description of how this system may have worked or even of the values assigned to the rows or columns. Wassén (1931: 198-199) has made an effort to infer this information by assigning the rows values of the exponents of 10 (starting with 1 at the bottom) and by assigning the values 1, 5, 15, and 30 to the columns (values which were multiplied by the row-value), but he does so solely on structural grounds, and not entirely convincingly. Nevertheless, it is unlikely that this board is a result of

[2] It is possible, though unproven, that some quipus may have recorded ideas or speech through some sort of numerical code. Since no key exists for such a code, we have only ascertainable numerical values with which to interpret quipus.

diffusion from Spain, since no comparable board was used in the 16th century anywhere in Europe (Wassén 1931: 204).

After the Spanish conquest in 1532, quipus continued to be used for the same administrative functions as they had been previously, and the data recorded on them were used by the Spanish (through Andeans who could read their values) to assist in collecting taxes and taking censuses (Loza 1998; Fossa 2000). The widespread use of quipus was eliminated in the 1580s, when they were declared to be idolatrous and the Spanish colonial administrators decreed that they should be destroyed. Nevertheless, quipus continued to be used among indigenous animal herders in parts of Peru and Bolivia for recording quantities of livestock (Bennett 1963: 618-19; Ifrah 1998: 69-70; Urton 1998: 410). The technique of recording was slightly different, in that often the pendant cords were simply tied together rather than using a main cord, and only single knots were used in place of the three-knot Inka system (Bennett 1963: 619). These were not simply "tally-knot" systems, however, but were cumulative-positional and decimal, and thus constituted a survival of the Inka numerical notation system. The extent of their present use seems very limited, if in fact they are used at all.

Obɛri Ɔkaimɛ

In the late 1920s, a syncretic indigenous-Christian religious movement known as Obɛri Ɔkaimɛ arose among a group of speakers of Ibibio-Efik in southeastern Nigeria. By 1931, the divinely inspired leaders of this movement had developed an alphabet (written from left to right) and a set of numeral symbols (Adams 1947; Hau 1961). The script was used for writing an arcane revealed liturgical language of the sect, but not for Ibibio-Efik. The Obɛri Ɔkaimɛ numeral-signs are shown in Table 10.2 (Hau 1961: 295).

Table 10.2: Obɛri Ɔkaimɛ numerals

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	0

The system is ciphered-positional and vigesimal; it is the only known system that is ciphered-positional and base-20 with no sub-base (with the possible partial exception of the Maya head-glyph numerals). Numeral-phrases are written from left to right with the highest exponents on the left. Thus, 1938 would be written as (4 x 400 + 16 x 20 + 18).

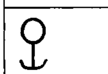
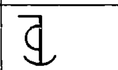
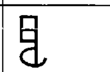
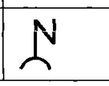
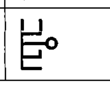
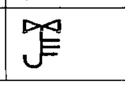
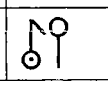
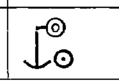
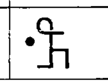
The inventors of the Obɛri Ɔkaimɛ numerals were educated in Christian missionary schools in the 1920s, where they became literate in English and learned Western numerals. While none of the numeral-signs has any graphic resemblance with the corresponding Western numerals except for 0, the script and its numerals were strongly influenced by Western traditions of writing, perhaps more than was any other indigenous African script (Dalby 1968: 160-161). Hau's (1967) highly dubious suggestion that the Obɛri Ɔkaimɛ script derives directly from Minoan Linear A, used thousands of kilometres away and over three millennia previously, cannot possibly apply to the numerals.

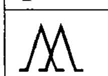
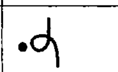
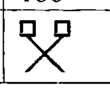
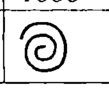
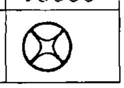
The numerals were used in a relatively small number of liturgical texts and personal letters among the members of the Obɛri Ɔkaimɛ sect. While the system was still used by some individuals in the 1950s, when Kathleen Hau (1961) corresponded with its leaders, its present use (if any) is not known. It is probably extinct. Western and sometimes Arabic positional numerals are used in the region today.

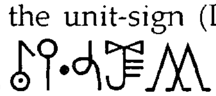
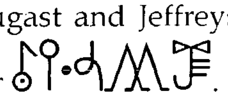
Bamum

The Bamum live in part of southwestern Cameroon near the border with Nigeria. In the early twentieth century (possibly around 1903[3]), Njoya, a Bamum ruler, took it upon himself to develop a script for his people. Njoya worked on his script incessantly until his death in 1933, starting with a large logosyllabary and gradually reducing the number of signs until he had created a syllabary of only 80 characters. From its inception, Bamum writing made use of numerical notation. The earliest Bamum numerals are shown in Table 10.3 (Dugast and Jeffreys 1950: 6).

Table 10.3: Bamum numerals (original)

1	2	3	4	5	6	7	8	9
								

1	10	100	1000	10000
				

This system is purely decimal and multiplicative-additive, with numeral-phrases written from left to right. Curiously, the exponent-sign for the units could either precede or follow the unit-sign (Dugast and Jeffreys 1950: 30). Thus, for instance, 76 could be written as  or . The unit-signs for 7, 8, 9, and 10 were not at this stage fully ideographic, but instead were constructed of two graphic parts, each of which represented a syllable in the two-syllable Bamum words corresponding to those numbers (Dugast and Jeffreys 1950: 98). In fact, at this point in the system's history, we may ask whether it is a numerical notation system or a set of lexical numerals. This is the same problem we encountered with the Shang / Zhou and Chinese classical systems (ch. 8), which not coincidentally also are multiplicative-additive and associated with logosyllabic scripts in which some characters (including numeral-signs) are ideograms.

[3] Dugast and Jeffreys (1950: 4) place its invention in 1895 or 1896, although other sources argue that Njoya did not develop a script until at least the turn of the century.

By 1921, the Bamum script had undergone several reductions and simplifications, but the numerals were still multiplicative-additive. Around that time, Njoya supervised the transformation of the script into a form known as *mfɛmfɛ*, at which time the system's structure was altered from multiplicative-additive to ciphered-positional by removing the exponent-signs (Dugast and Jeffreys 1950: 30). The old sign for 10 took over the role of zero, and numeral-phrases were written from left to right with digits for 0 through 9, just as the Western and Arabic positional numerals which were the Bamum system's primary rivals. The *mfɛmfɛ* numerals are shown in Table 10.4 (Dugast and Jeffreys 1950: 31).

Table 10.4: Bamum numerals (*mfɛmfɛ*)

1	2	3	4	5	6	7	8	9	0
ꛦ	ꛧ	ꛨ	ꛩ	ꛪ	ꛫ	ꛬ	ꛭ	ꛮ	ꛯ

During its heyday in the first three decades of the twentieth century, the Bamum numerals were used quite widely, no doubt primarily due to Njoya's political clout. The numerals were employed on a variety of legal documents, census records, histories, and personal letters, both handwritten and printed. Njoya was deposed in 1931 and died two years later, after which time the Bamum script and numerals rapidly fell into disuse. Nevertheless, the Bamum numerical notation systems, like the script, are more than just a historical curiosity because we are able to trace their rapid transformation from an additive to a positional structure by the simple step of removing the exponent-signs from numeral-phrases.

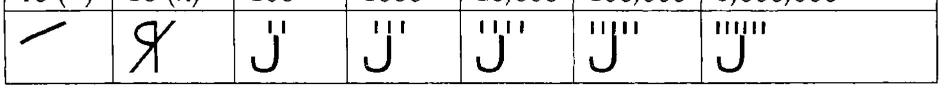
Mende

Just prior to 1920, a syllabary known as *Kikakui* was developed by a tailor, Kisimi Kamara, in order to represent graphically the Mende language spoken in Sierra Leone. In addition, Kisimi Kamara developed a numerical notation system to accompany his new invention. The Mende numeral-signs are shown in Table 10.5 (Tuchscherer 1996: 71-75).

Table 10.5: Mende (Kikakui) numerals

1	2	3	4	5	6	7	8	9
𞣇	𞣈	𞣉	𞣊	𞣋	𞣌	𞣍	𞣎	𞣏

10 (+)	10 (x)	100	1000	10,000	100,000	1,000,000



The system is decimal and multiplicative-additive, and numeral-phrases are constructed with the highest exponents on the right. Because the system is multiplicative-additive, no sign for zero is needed or used. Unit-signs are placed above the corresponding exponent-signs, and so numeral-phrases are read from top to bottom and from right to left. The system is slightly irregular. There are two signs that mean 10. The first, which I have identified as 10(+) in Table 10.5, is used additively in combination with the units for 1 through 9 in order to write 11 through 19 (Tuchscherer 1996: 72). The other, 10(x), is the standard multiplicative exponent-sign for 10 used in combination with the unit-signs for 2-9. The 10(x) sign is also used to indicate 10 alone by placing a dot rather than a sign for 1 above it (Tuchscherer 1996: 71). A notable feature of the higher exponent-signs is their graphic use of vertical strokes to indicate repeated multiplication by 10; the number of strokes represents the power of 10 corresponding to the number. This is quite distinct from the cumulative principle, which always refers to repeated *addition* of similar symbols, and it is a feature that is unique to the Mende system. In theory, it suggests that the system could have been extended infinitely without using the positional principle, although there are practical limits to how many vertical-strokes could be read and used easily. Figure 10.3 indicates a selection of Mende numeral-phrases.

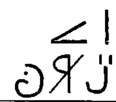
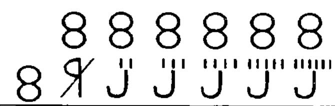
14	
128	
60,009	
5,555,555	

Figure 10.3: Mende numeral-phrases

It was originally thought by some researchers that the numeral-signs for 1 through 10 were derived acrophonically from the *Kikakui* signs corresponding to the first syllables of the numeral words for 1 through 10 (Tuchscherer 1996: 130-132). While there is an exact correlation between these syllabic values and the numeral-signs, this does not prove the validity of this causal path. Rather, Karl Tuchscherer (1996: 140-142) has demonstrated that the Mende numeral-signs (at least those for 1 through 5) are similar to certain signs (and variants) of the Arabic positional numeral-signs. From this, he argues that the Arabic numerals helped inspire some of the signs of the *Kikakui* syllabary (at least those for the first syllables of number words) rather than vice versa. While, to my eye, the similarities are not striking enough to prove the case conclusively, I am reasonably convinced that the Arabic positional numerals are more likely than any other system of the region to have influenced the development of the Mende system. Yet the Mende numerals are multiplicative-additive, not ciphered-additive (like the abjad-derived systems of the Muslim world) or ciphered-positional (like most of the other systems used in the region). The only other multiplicative-additive system used in West Africa is the earliest Bamum system, but it is a long way from Sierra Leone to Cameroon, and by the time the Mende system was developed in 1921, the Bamum had switched to ciphered-positional numerals. Moreover, the use of two different signs for 10 (one additive, one multiplicative) and the use of repeated strokes to indicate successive multiplication by 10 are features that are not attested in other possible ancestral systems. Thus, the structure of the Mende system should be regarded as largely if not wholly indigenous. Curiously, the modern Mende lexical numerals are not decimal but vigesimal. While this might suggest that the base of the Mende numerical notation was borrowed from the Arabic numerals, in the nineteenth century the Mende had decimal lexical numerals (Tuchscherer 1996: 148-150). If this system survived (in even a vestigial form) into the first decades of the twentieth century, it, rather than a foreign numerical notation system, could have been the inspiration for the decimal base of the system.

The Mende numerals were used for a wide variety of functions, and were taught in schools throughout the 1920s and 1930s. Some individuals used the system for accounting and record keeping, but it is not clear whether the numerals themselves were used directly for arithmetic (Tuchscherer 1996: 69). Dalby reports that the syllabary was used by some weavers and carpenters for recording measurements, which would presumably also require numerals (Dalby 1967: 21). Today, the Mende numerals are essentially extinct, and Western or Arabic numerals are used for all functions.

Sub-Saharan decimal-positional

In addition to the African systems described above, which are structurally distinct from their ancestors, several of the indigenous scripts of sub-Saharan Africa have numerical notation systems that are basically decimal and ciphered-positional, and are thus structurally identical to their Western or Arabic ancestors. While these systems are of less interest from a structural point of view, they are noteworthy from a historical perspective, not least because most of them have not been mentioned in other histories of numeration. To remedy this deficiency, I list these systems in Table 10.6.

Table 10.6: Decimal systems of sub-Saharan Africa

	1	2	3	4	5	6	7	8	9	0	10
Bagam											
Bété										●	
Fula (Dita)											
Fula (Adama Ba)											
Kpelle											
Manding										0	
Wolof	|									○	

The Bagam script was a syllabary invented early in the twentieth century in western Cameroon and used by the Eghap (known in scholarly literature as the Bagam) of that region for a brief period (Tuchscherer 1999). The only text to preserve Bagam writing and numerals is a recently discovered 1917 description of the system by a British colonial military officer, Captain L.W.G. Malcolm. The numerals probably were derived from the Bamum system rather than the Western numerals. A few graphic resemblances can be seen between the Bamum and Bagam sign sets. The Bagam numerals do not include a sign for zero, but do include a sign for ten. It is thus unclear whether it was a ciphered-positional system or how (if at all) it expressed higher numbers. In the early part of the century, the Bamum system was still multiplicative-additive, which is suggestive but not conclusive that the Bagam system may also have had this structure. The Bagam script and numerals are now extinct, and recent ethnographic investigations in the region revealed no knowledge of the numerals even among elderly Bagam (Karl Tuchscherer, personal communication).

The Bété numerals were invented in late 1957 or early 1958 by Frédéric Bruly-Bouabré, a native Bété from the western part of Ivory Coast, to accompany a syllabary of over 400 characters that he had invented a year or so earlier (Monod 1958). Bruly-Bouabré, who was fully literate in French, did not use Western models in developing his script-signs, as can be seen from the abstract nature of the numerals. However, the use of a dot for 0 shows at least some influence from the Western numerals (or perhaps the Arabic numerals, although it is not clear whether Bruly-Bouabré knew Arabic at all). The role of the sign for 10 is unclear and it is unknown whether it was used multiplicatively or additively in conjunction with the unit-signs. There is evidence of a quinary component to the Bété system in the fact that the signs for 6 through 10 are inverted forms of the signs for 1 through 5, with the exception of the extra dot atop the sign for 5 (Monod 1958: 437). Bruly-Bouabré's efforts to have this system accepted among the Bété appears to have met with minimal success. I do not know whether it is still used at present.

Two alphabets invented for the Fula of Mali have accompanying ciphered-positional decimal numerical notation systems. The first of these, known as *Dita*, was developed by Oumar Dembéle between 1958 and 1966; in keeping with his being a woodworker, his signs have a linear character (Dalby 1969: 168-173). Dembélé attended a Koranic school and spoke French, so the structure of the system was based either on Western or Arabic numerals. The second system, invented by Adama Ba, a Fula Muslim literate in French, before 1964, is identical in structure but its signs are more curvilinear and perhaps show some influence from Western numerals (Dalby 1969: 173-174). Neither of these two systems was ever used except by their inventors.

The Kpelle numerals were developed in the 1930s by Gbili, a paramount chief of the Kpelle in central Liberia, in conjunction with an indigenous syllabary (Stone 1990). Its numeral-signs include a sign for 10 but none for zero, so it is not clear how, if at all, higher numbers were written. Both Arabic and Western numerals were known in the region, and either could have been the inspiration for the Kpelle system, but, although the Kpelle signs show vague graphic resemblances with both Arabic and Western numerals, no definite origin can be assigned to them. It is possible that the signs are entirely indigenous in origin. The script was used traditionally for tax records as well as for official communication among chiefs, but seems to have been restricted to a small segment of the populace. Today, most Kpelle use Western numerals, and the indigenous system is known by very few individuals (Stone 1990: 141).

A set of numerals was developed around 1950 by Souleymane Kantè, an educated trader who was literate in both French and Arabic, in conjunction with an alphabet known as *N'ko* (Dalby 1969: 162-165). It was designed for use among the many peoples whose dialects fall under the label "Manding", most notably Mandinka, and was intended to provide a means of communication accessible without the need for formal schooling. The numerals are ciphered-positional and decimal, and perhaps are related graphically to the Western numerals, but contrast with both the Western and Arabic systems in that numeral-phrases are written with the highest exponent on the right. Texts written in this script apparently included treatises on calculation, suggesting that the numerals may have been used for arithmetic (Dalby 1969: 163). *N'ko* continues to be used today, and probably has tens of thousands of users.

Assane Faye developed a Wolof script around 1961 that has a set of ciphered-positional numerals (Dalby 1969: 165-168). Faye, who was literate in both French and Arabic, presumably drew more influence from the Western numerals in creating this system, whose signs show a stronger graphic resemblance to Western than to Arabic numerals. Numeral-phrases were written from left to right. Curiously, Faye also assigned numerical values to nineteen of the letters of his script (1-9, 10-90, 100) in imitation of the Arabic abjad ciphered-additive system (Dalby 1969: 167-168). Neither the script nor the numerals survives today; most Wolof use either Arabic or Western numerals.

Pre-colonial West Africa

Often, when researchers discuss the history of post-colonial scripts and numerals of West Africa, it is to portray them as derivative and thus to denigrate the inventiveness of Africans. While some of these systems (e.g. Mende, Bamum, Obɛri Ɔkaimɛ) are structurally distinct from the Western and Arabic numerals, I grant that these systems probably would not have developed without colonialism and contact with the West. At the same time, however, there is suggestive evidence that pre-colonial West Africans also used numerical notation. Unfortunately, we do not have a full understanding of the numerical notation systems used in West Africa prior to the colonial period. Instead, we have a handful of ethnographic details pertaining to the peoples of West Africa in the twentieth century concerning systems that may be considerably older. While we should not assume that these systems are of entirely indigenous origin, given that contact with Muslim traders from the north could have provided a powerful stimulus towards inventing such systems, neither should we discount the possibility. Even among historians of mathematics interested in African capabilities, these systems have not been discussed, doubtless because the researchers in question were unaware of them (Zaslavsky 1973, Gerdes 1994). Because these systems are not attached to phonetic scripts, they have not been compared to other numerical notation systems. A further problem is that true numerical notation systems (structured intra- and interexponentially and having a base) are often conflated with unstructured tally-systems. I fully expect that a more thorough search of the relevant ethnographic literature (especially from the early twentieth century) would reveal additional numerical notation systems.

A.S. Judd (1917), reporting on the state of education in northern Nigeria, reported that a numerical notation system was employed by the Munshi people. This system, which has "a thin line representing the units, a circle the tens, and a broad line made by the thumb representing a score", was apparently used when drawing in sand or earth

(Judd 1917: 5). Presuming that Judd's description is accurate, the most likely possibility is that this system was cumulative-additive with a base of 20 and a sub-base of 10.

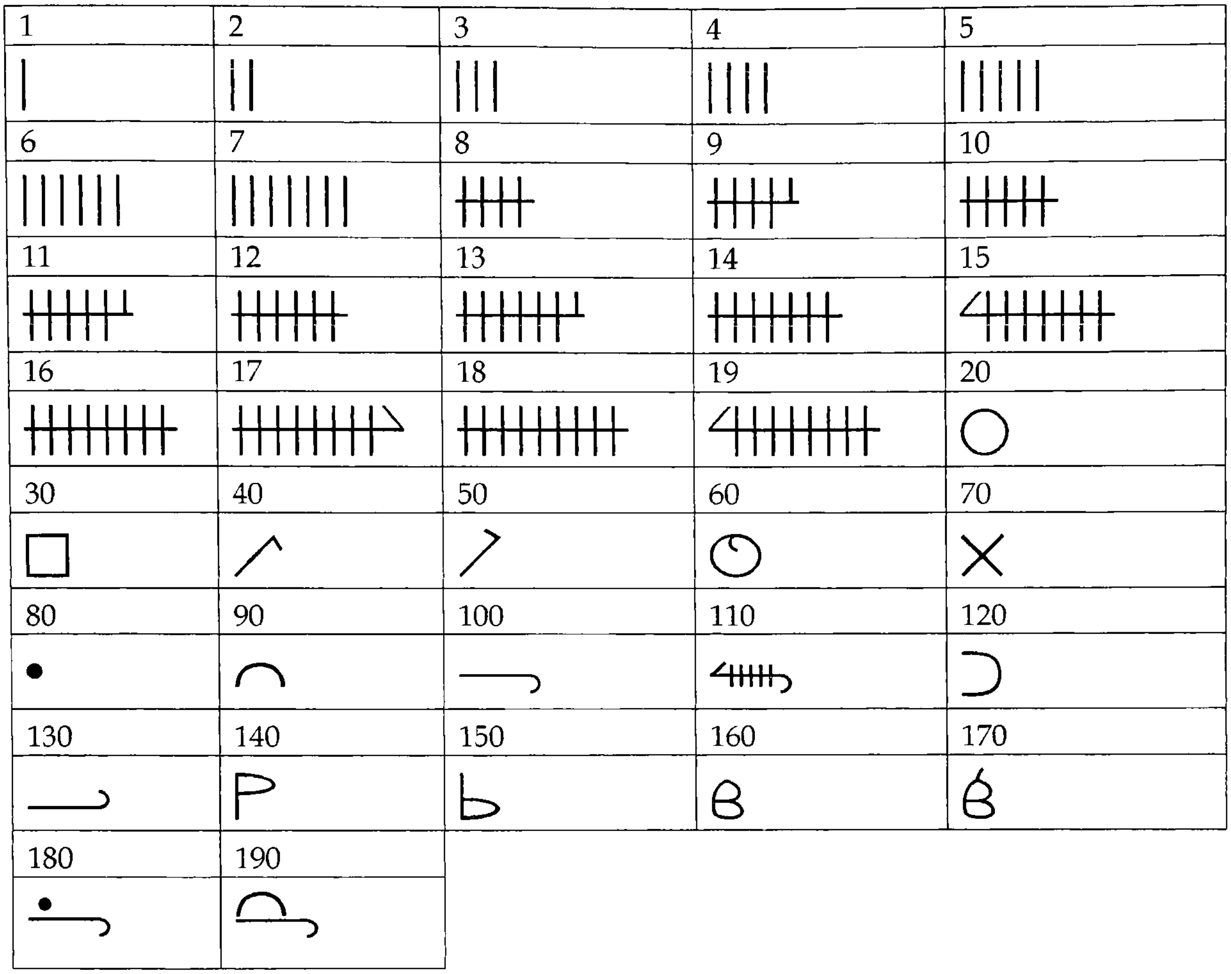
The tradition of graphic symbolism practiced by the Dogon of Mali in rock paintings and sand drawings includes numerical signs that can be combined with one another. In some symbols, straight lines represent units and circles represent five; a drawing of a man with four circles (each representing one of the limbs with five digits) joined with a cross carries the numerical significance 22 (Griaule and Dieterlen 1951: 11-12; Flam 1976: 37). Another symbol represents a period of sixty years by three rods of decreasing size, each with the value of 20 (Griaule and Dieterlen 1951: 28). It seems that, at least at the time the researchers were present, there may not have been a regular system of correspondences between numbers and signs. In the context of reckoning and calculation, cowries representing 1, 5, 10, 20, 40, and 80 apparently were used (Calame-Griaule 1986: 232). The exact technique employed is unknown, however, and this may not have constituted a numerical notation system either.

While most systems of tally-sticks use only one-to-one correspondence, Lagercrantz (1973: 572) reports that among the Ganda and Djaga, tally-sticks are also used in which units are marked by small notches, 10 by a larger notch and 100 by an even larger notch. It is not clear whether this system is used for recording cardinal numbers, or whether it is simply a series of marks equal to the number being counted, of which the tenth is large and the hundredth larger still.

Bambara

One of the most peculiar African numerical notation systems was used by the Bambara of Mali in religious and divinatory contexts (Ganay 1950). Although details of the system's history are rather sketchy, we do have a fair idea of the numeral-signs and the structure of the system. The Bambara numeral-signs are shown in Table 10.7.

Table 10.7: Bambara numeral-signs (Ganay 1950: 302-305)



The Bambara system is quite irregular structurally; while it is additive throughout, it alternates between cumulative and ciphered notation, and while it is mainly decimal, it has vigesimal components. For instance, 1 to 19 are written primarily with vertical cumulative unit-strokes. The value of a set of vertical strokes is doubled if a horizontal line is crossed through it (effectively dividing the number into two registers, one above and one below the line). For odd numbers, an additional half-stroke can be placed at either end of the phrase, sometimes vertically and other times at an angle. Each of the tens from 20 to 170 has its own sign, which makes the system ciphered at this point. The signs for 180 and 190 are additive combinations of 100+80 and 100+90, respectively. To add a number of units from 1 to 9 to one of these ciphered signs, an appropriate number of strokes are attached to the sign for the multiple of 10 (or dots, when adding units to 60, 160, or 170). Thus,  represents 68 and  represents 189. This means of representation is decimal, insofar as each decade has its own sign to which up to nine unit-signs could be attached. Yet, because there are signs for 110, 120, and so on, it is not a ciphered-additive decimal system like the Greek alphabetic numerals (in which 100 is followed by 200, 300, and so on). Moreover, some of the decade-signs are similar enough to the ones preceding them (40 vs. 50, 100 vs. 110, 140 vs. 150, 160 vs. 170) to suggest an additional trace of a vigesimal base. For numbers higher than 200, the cumulative principle is again employed by repeating the sign for 100 (another decimal component) as many times as required in a vertical column, with any needed additional signs placed at the top of the column. Figure 10.4 shows some higher numeral-phrases (as reproduced from Ganay 1950: 300).[4]

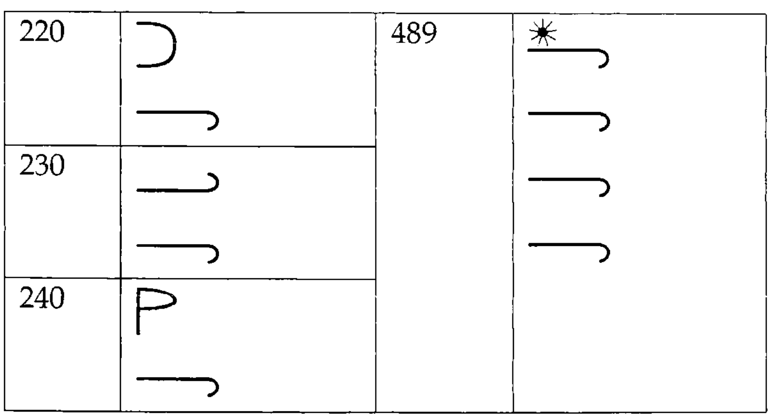


Figure 10.4: Bambara numeral-phrases

The Bambara numerical notation system seems to have been used primarily in ritual contexts, especially those pertaining to divination using numbers (Ganay 1950: 298). Nothing can be said of the origin, period of use, or decline of this system. It shows no resemblance to any of the systems that would have been known by Bambara, who had considerable contact with the Muslim world. While the Arabic abjad numerals commonly used for divination in the Maghreb seem the most likely ancestor, and indeed they are

[4] Large numeral-phrases for 1935 and 4000 are also listed, but are highly irregular, and I cannot determine what principle has been used to determine their value.

ciphered-additive, in all other respects - its frequent use of the cumulative principle, the presence of a vigesimal component, and its numeral-signs - the Bambara system is quite different. I consider it very likely to have been indigenously invented. I have no idea whether this system continues to be used, though I suspect that it is not.

Varang Kshiti

In the twentieth century, several scripts were developed for the various Munda languages of central and eastern India, of which Sorang Sompeng, Ol Cemet', and Varang Kshiti are the primary ones to survive to the present day (Zide 1996). While these scripts have numerical notation systems, most are ciphered-positional and thus are clearly derived from the Western numerals or the ciphered-positional systems of India (ch. 6). I know of only one script, the Varang Kshiti script designed for the Ho of Bihar province, where a structurally distinct numerical notation system was developed for a Munda language. These numerals are shown in Table 10.8.

Table 10.8: Varang Kshiti numerals (Pinnow 1972: 828)

1	2	3	4	5	6	7	8	9
𑣡	𑣢	𑣣	𑣤	𑣥	𑣦	𑣧	𑣨	𑣩
10	20	30	40	50	60	70	80	90
𑣪	𑣫	𑣬	𑣭	𑣮	𑣯	𑣰	𑣱	𑣲

The system appears to be ciphered-additive, as it has signs for 1-9 and 10-90. The signs for 10 through 30 do not resemble the signs for the corresponding units, but the higher decades obviously do. Curiously, however, Pinnow (1972: 830) reports that only the unit-signs, combined in a ciphered-positional manner, were employed when writing numbers from 11-19, 21-29, 31-39, etc. The separate signs for the decades from 10 to 90 may have served to obviate the need to introduce a sign for zero in order to write any number less than 100. If this analysis is correct, then 60 would have been written simply as  but 61 as . There may also have been signs for 100 and 1000, which presumably would combine multiplicatively with the unit-signs, but this cannot be confirmed (Pinnow 1972: 831).

The Varang Kshiti script and numerical notation system were developed by a Ho shaman named Lako Bodra throughout the 1950s and 1960s. While various claims have been made concerning the antiquity of the script (such as that it was first developed in the 13th century and rediscovered by Lako Bodra in a vision), it is likely that it is a recent invention (Zide 1996: 616-617). This fact does not help us determine upon what model, if any, the numerical notation system was based. There are similarities in the numeral-signs with those of various South Asian systems, but none of these is suggestive enough to prove a specific origin. Pinnow (1972) believes at least some of the script-signs to have been borrowed from ancient Brāhmī characters. Since both Varang Kshiti and Brāhmī numerical notation systems are ciphered-additive, I do not discount this possibility entirely, but there is no evidence that the Varang Kshiti system is of sufficient antiquity to have been influenced by Brāhmī.

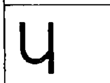
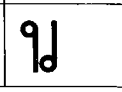
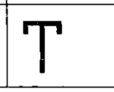
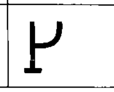
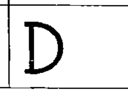
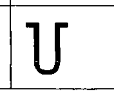
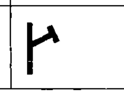
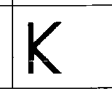
The Varang Kshiti script and numerals are still used in both primary and adult education, and efforts to make it the vehicle for strengthening Ho culture have met with some success. I strongly suspect that in most circumstances, Western, Devanagari, or Oriya numerals are used in place of the system described above.

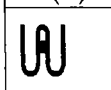
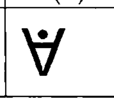
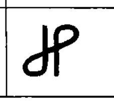
<u>Pahawh Hmong</u>

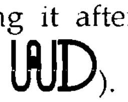
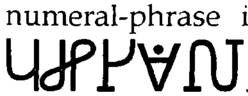
The Pahawh Hmong script was developed for speakers of the Hmong language of northern Laos. Its inventor, a Hmong peasant named Shong Lue Yang, though apparently illiterate when he developed this script, revised it constantly from 1959 until his assassination in 1971 and used it as a tool to promote Hmong cultural identity (Ratliff

1996). In addition to phonetic script-signs, Shong Lue Yang and his disciples developed a numerical notation system. The earliest (Source Version) of the Pahawh Hmong numerals are shown in Table 10.9 (Smalley 1990: 79).

Table 10.9: Pahawh Hmong (Source Version) numerals

1	2	3	4	5	6	7	8	9
								

10(+)	20	10(x)	100	1000
				

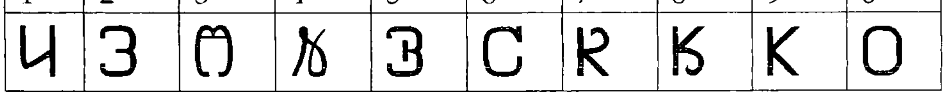
This system is primarily multiplicative-additive and decimal; unit-signs from 1 through 9 combine with exponent signs for 10, 100, and 1000. Numeral-phrases, like the script itself, are written from left to right. The only irregularity in the system is that 10 and 20 are not expressed through multiplication of the unit-signs 1 and 2 with the exponent-sign for 10, but with distinct signs, which also are combined additively with the unit-signs to write 11-19 and 21-29. The other sign for 10, shown as 10(x) in Table 10.9, is used multiplicatively by placing it after the unit-signs 3-9. Thus, 36 is written as  while 16 is written as ). The use of two signs for 10 (one additive and one multiplicative) is parallel to the multiplicative-additive Mende system described earlier in this chapter. While there are no Pahawh Hmong numeral-signs for 10,000 or higher exponents, these numbers could be written multiplicatively by placing an entire numeral-phrase in front of the sign for 1,000; for example 150,000 is written as .

The Source Version Pahawh Hmong script was developed around 1959, but since Shong Lue Yang was apparently illiterate at the time, it is difficult to say whether some other numerical notation system had any influence in its invention. Given that the standard Chinese numerals are multiplicative-additive, it seems possible that they played some part in the origin of the Pahawh Hmong system. This is supported by the use of a special sign for 20 (廿) in Chinese, though Pahawh Hmong, unlike the Chinese system, does not have distinct signs for 30 and 40. At any rate, the Pahawh Hmong numeral-signs are entirely different from the Chinese ones, so I think it is best not to presume any influence from China.

Within about ten years, Shong Lue Yang and his followers developed a new system based in part on the old numeral-signs. By this period, the script used was that known as the Second Stage Reduced Version. The numerals of this period are shown in Table 10.10 (Smalley 1990: 80).

Table 10.10: Pahawh Hmong (Second Stage Reduced Version) numerals

1	2	3	4	5	6	7	8	9	0



This is obviously a ciphered-positional, decimal system. Some of the numeral-signs from the Source Version are similar or identical to the ones in this system (1, 3, 9), but many others are changed entirely. The addition of a sign for 0 and the abandonment of the exponent-signs change the system's structure radically. It is probable that this transformation was a result of a growing awareness of Lao and/or Western numerals by Shong Lue Yang, although the numeral-signs (excepting the zero) are unlike those of any neighbouring systems.

Despite the adoption of ciphered-positional numerals in the Second Stage Reduced Version, multiplicative-additive notation was not abandoned but was in fact expanded by Shong Lue Yang. A new Pahawh Hmong multiplicative-additive system was used alongside the ciphered-positional system, that combined the unit-signs for 1-9 from the Second Stage Reduced Version with a new set of exponent-signs. These signs are shown in Table 10.11 (Smalley 1990: 81).

Table 10.11: Pahawh Hmong Second Stage Reduced Version exponent-signs

10	100	1000	10,000	100,000	1,000,000	10,000,000	100,000,000
𖭛	𖭜	𖭜𖭛	𖭝	𖭝𖭛	𖭞	𖭞𖭛	𖭟
1,000,000,000	10,000,000,000	100,000,000,000	1,000,000,000,000				
𖭟𖭛	𖭠	𖭠𖭛	𖭡				

Rather than creating a separate exponent-sign for each exponent of 10, Shong Lue Yang hit on the idea of using distinct signs only for the exponents of 100 (100, 10,000, 1,000,000, etc.), using the exponent-sign for 10 multiplicatively with these signs to write the intermediate exponents (1000, 100,000, etc.). This cut in half the number of new exponent-signs that needed to be invented. Because of this additional structural element, this form of Pahawh Hmong numeration, while still multiplicative-additive, is not simply decimal; since exponents of 100 (not just exponents of 10) structure the system, it is also centesimal. As Smalley (1990: 81-82) points out, there is a considerable advantage in conciseness to be gained when writing large round numbers in this system as compared to in the ciphered-positional one.

Both the ciphered-positional and the revised multiplicative-additive Pahawh Hmong systems continue to be used, although only the ciphered-positional system is used for arithmetical calculation. I do not know with what frequency or in what contexts each of these systems is used. Because large numbers of Hmong have immigrated to the West (especially Australia), Hmong numerals are used not only in Laos and Hmong-speaking parts of Vietnam, but also in Western countries. At present, the Hmong numerals are challenged by both the Lao and Western ciphered-positional numerals, so it is not clear how long they will continue to be used. Regardless of the eventual success of any of these systems, it is noteworthy that three variants of the Pahawh Hmong system were developed with such rapidity, each with a different structure.

Zuñi

There is no evidence for numerical notation in the New World north of Mexico prior to the European conquest. Yet a single object described by the renowned nineteenth-century ethnographer, Frank Hamilton Cushing, suggests that, at least in the 1890s, the Zuñi of the American Southwest used a decimal cumulative-additive numerical notation system with a base of 10 and a sub-base of 5 (Cushing 1892). Figure 10.5 is a depiction of what Cushing (1892: 300) calls an "irrigation tally stick".

Figure 10.5: Zuñi irrigation tally

On the right side of this object, reading from right to left, there are 24 marks, of which the fifth and fifteenth are marked with a slanted stroke, and the tenth and twentieth are marked with an X.[5] On the left side, reading from left to right, there are two X marks, a vertical stroke, and a slanted line, which is amenable to the interpretation of 24 if a subtractive component to the system is assumed, as Cushing does (1892: 298). While the right side is a simple tally (it is grouped, but does not reduce multiple signs to a single one), the left side is a cumulative-additive numerical notation system in which I represents 1, \ represents 5, and X represents 10.

In addition to this system, Cushing reports the use of a system of knot-numerals (Cushing 1892: 300-302). Like the tally-stick system, it is cumulative-additive with a base of 10, a sub-base of 5, and uses subtractive notation for both 4 and 9. It uses a single knot for the units 1 through 3, a more complex knot - known as a 'thumb-knot' - for 5 and an even more complex 'double thumb-knot' for 10. These were combined in a cumulative-additive fashion, with 4 and 9 denoted by placing a single knot in front of a thumb-knot

[5] To be precise, the two slanted notches differ slightly, as do the two X marks, but it is not clear from Cushing's drawing exactly what the distinctions are.

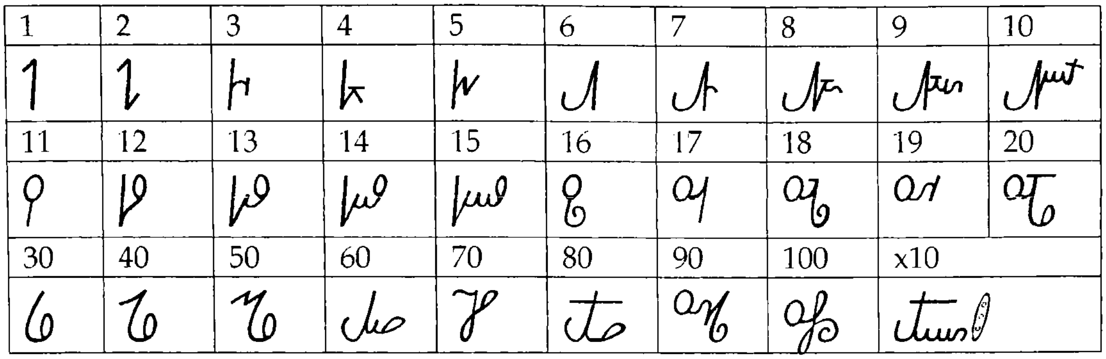
or double thumb-knot, respectively. While Cushing calls these knots 'quippos' and finds them to be parallel to the cumulative-positional Inka numerals, the Zuñi system is entirely additive, and has the additional features of a quinary sub-base and a subtractive component.

I do not know how extensively the knot and tally numerals were used among the Zuñi, or whether they were used for other functions. The irrigation stick is strikingly similar to tally-sticks used by Europeans, and both the tally and knot-numerals are essentially identical to Roman numerals (both are cumulative-additive, have a base of 10 with a sub-base of 5, and use subtraction for 4). Roman numerals would have been used by the early missionaries in the Southwest who worked among the Zuñi. If this is an independent invention, it parallels the Roman numerals and is even more striking than the parallel I described in Chapter 8 between the Roman and Ryukyu numerals. Given the extent of the similarity, it seems entirely plausible that this means of representation might have been borrowed from European sources. Nevertheless, I do not discount the hypothesis that it may have been an indigenous invention.

Cherokee

One of the most famous instances of stimulus diffusion, in the form of the indigenous invention of a script by an illiterate person on the basis of hearsay knowledge, was the creation of a syllabary for the Cherokee (Tsalagi) language around 1820 by Sequoyah. It is less commonly known that several years after inventing his syllabary, probably around 1830, Sequoyah also developed a decimal numerical notation system. The numerals of this system are shown in Table 10.12 (Holmes and Smith 1977: Appendix II and III).

Table 10.12: Cherokee numerals



1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
30	40	50	60	70	80	90	100	x10	

The system is ciphered-additive for numbers from 1 to 99. There are distinct signs for 1 through 19 and 20-90. The signs for 1-20 seem to be grouped graphically into sets of 5 (1-5, 6-10, 11-15, 16-20), but there is no simple relation (such as through the use of cumulative strokes) among the signs in each sub-grouping. This vigesimal element is very curious, since the Cherokee lexical numerals are purely decimal. Presumably, the signs for the tens between 20 and 90 combine additively with the unit-signs for 1-9, while the signs for 10 through 19 are used only on their own. The documentary evidence neither confirms nor refutes this supposition, but if the signs for 10 through 19 were combined with the signs for the tens, the signs for 30, 50, 70, and 90 would have been redundant. For writing numbers above 100, the system is not ciphered-additive but multiplicative-additive, and thus it is a hybrid system.[6] The sign indicated as "x 10" in the table always combines multiplicatively with the sign for 100, and multiplies the value of the phrase by ten. Perhaps I am over-interpreting the first element of this sign, but it strikes me as being similar to the cursive English word 'times'.

While Sequoyah had hoped that his numerical notation system would be adopted, just as the syllabary had, when he laid it before the Cherokee tribal council, they voted against it and in favour of the Western numerals (Holmes and Smith 1977: 293). As a result, we know of the Cherokee system from only two documents, both in Sequoyah's

[6] Strikingly, the Cherokee system is structurally identical to the Jurchin system (ch. 8), including the use of distinct signs for 10-19, and is very similar to the Ethiopic (ch. 5) and Sinhalese (ch. 6) systems, both of which are ciphered-additive below 100 and multiplicative-additive above.

hand, and only one of which transliterates the numeral-signs into Western numerals. One of these documents, prepared by Sequoyah for his friend John Howard Payne, is dated 1839 (in Western numerals) by Payne, suggesting that Sequoyah may still at that time have been attempting to resuscitate his system's fortunes (Holmes and Smith 1977: Appendix III). I do not believe that anyone ever used this system after Sequoyah's death in 1843, although some modern Cherokee appear to be aware of this system and its structure.

<u>Iñupiaq</u>

The newest numerical notation system is, at the time of writing, scarcely half a dozen years old. It was devised in 1995 by a group of Inupiat youth in Kaktovik, Alaska (located on Alaska's Arctic coast about 100 km from the Alaska-Yukon border) as part of a middle-school classroom project, and has been adopted more widely among the Inupiat.[7] The numeral-signs of this system are shown in Table 10.13.

Table 10.13: Iñupiaq numerals

1	2	3	4	5	6	7	8	9	10
𝋁	𝋂	𝋃	𝋄	𝋅	𝋆	𝋇	𝋈	𝋉	𝋊
11	12	13	14	15	16	17	18	19	0
𝋋	𝋌	𝋍	𝋎	𝋏	𝋐	𝋑	𝋒	𝋓	𝋀

The system is cumulative-positional with a base of 20 and a sub-base of 5. The numeral-signs are written using slightly diagonal vertical strokes with a value of 1, above which slightly diagonal horizontal strokes are placed, each with a value of 5. When the numerals are handwritten, the vertical and horizontal strokes are of the same width, but

[7] My information on this system is based entirely on very fruitful discussions with W. Clark Bartley, the non-Inupiat instructor of the mathematics class in which the system was developed.

sometimes in print the horizontal strokes are shown somewhat thicker than the vertical strokes. The zero-sign is reported to be graphically symbolic of a human figure's arms crossed over the chest, but is also similar to the Western zero-sign.

We are fortunate to have enormous detail regarding the circumstances and thought processes of the Inupiat inventors of the system. The students, having completed work on binary notation, realized that the lexical numerals of the Iñupiaq language were base-20, and took it upon themselves to develop a vigesimal numerical notation system that would better correspond with their lexical numerals.[8] At first, an attempt was made to develop ciphered signs for 10 through 19, but this was found to be taxing on the memory of users. The students turned instead to a cumulative-positional system that requires only two different strokes (vertical for ones, horizontal for fives) and a zero. At the time, neither they nor their teacher were familiar with other cumulative-positional numerical notation systems such as the Chinese rod-numerals (ch. 8) or the quasi-positional Mesoamerican bar and dot numerals (ch. 9). The numeration tools possessed by the students comprised the Western numerals as well as a brief introduction to Chisanbop finger-computation, a quinary-decimal calculating technology that also may have inspired them.

The Iñupiaq system is unusual among numerical notation systems in that its invention was specifically in the context of mathematical education; its design was always meant to aid students in working with arithmetic. Although the choice of cumulative-positional notation with a sub-base seems to have been stimulated by the difficulty entailed in memorizing 20 separate symbols that a ciphered system would have required, it had the added effect of facilitating arithmetic using physical counters. Techniques were quickly developed to manipulate numbers using popsicle sticks to

---

[8] While the Iñupiaq lexical numerals are vigesimal with a sub-base of 5, they deviate from the numerical notation system described here in the use of subtractive formations for 9 (10-1), 14 (15-1), and 19 (20-1), as well as in the use of a word for 6 that is not derived from that for 5.

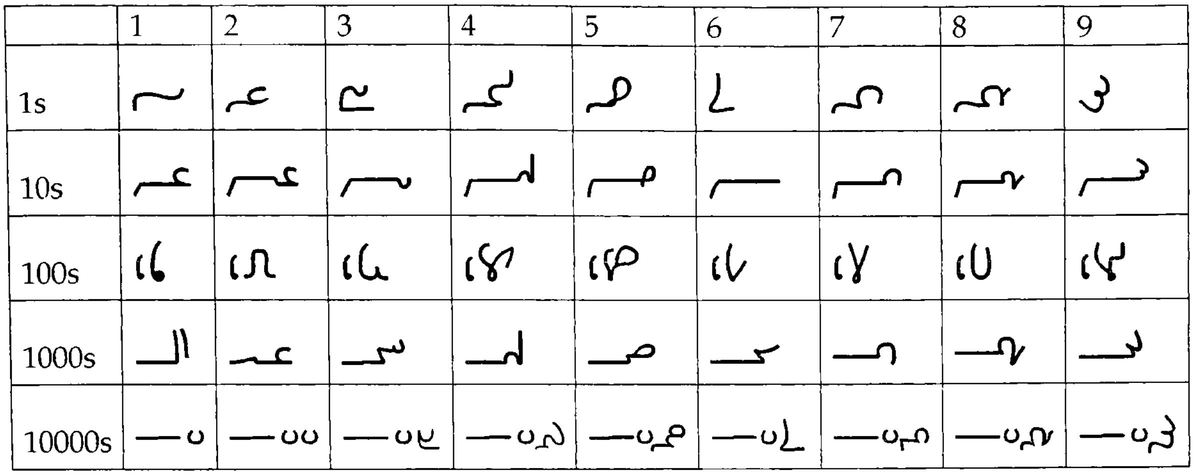
represent the vertical and horizontal strokes of the written numerals, thus producing a computational device whose results could be written on paper easily thereafter. In some cases, the students found it more convenient to use this device in a purely base-5 fashion (i.e. with up to four horizontal sticks for 5 instead of only three, as in the numerical notation system).

While the system is still understood by a number of youth of northern Alaska, as well as by some educators, its eventual success is still very uncertain, as Western numerals are strongly preferred by many educators. Although Inupiat children trained in this system have had considerable success in their mathematics education, the very small number of users of this system limits its present value as a communication tool. It is too early to say whether the official adoption of this system by the Commission on Inupiat History, Language, and Culture will help its chances of survival.

Siyaq

A very unusual set of numerical notation systems was employed by Arab, Persian, Islamic Indian, and Ottoman administrators between the 10th and 19th centuries for representing numbers in financial transactions. While they are known by many names (*dewani* by the Arabs, *siyaq* by the Persians and Turks, and *rokoum* in India) and exhibit enormous paleographic variability, they all share a common origin and structure. Despite these systems' use in several important states over nearly a millennium, they have been ignored by most Western scholars. Recognizing that it is slightly inappropriate to refer to all variants of the numerals as 'siyaq', I will group them all here under this single term. The Persian siyaq numeral-signs are shown in Table 10.14 (Kazem-zadeh 1915: Plates I-III).

Table 10.14: Siyaq numerals



The siyaq numerals have nine distinct signs for each exponent of 10, and thus the system is basically ciphered-additive and decimal. Numeral-phrases are written from right to left, although in numeral-phrases containing both units and tens, the unit-sign is found to the right of the tens-sign (i.e. before it rather than after it). Often, individual signs are ligatured together, making it difficult in some cases to distinguish the individual components of a numeral-phrase. There are similarities among the signs for different multiples of the same exponent. For instance, the signs for the tens all have a short diagonal stroke connected to a long horizontal stroke ( ), followed by some additional component, whereas the signs for the hundreds all have a small curved line at the left followed by some other component. Moreover, there are similarities among the signs for the same multiples of different exponents. Thus, 9, 90, 9000, and 90000 all have the common element (900 is a special case, and does not conform to this rule). These similarities suggest that we might try to understand the system as a multiplicative-additive one in which each sign is composed of a unit-sign on the right and an exponent-sign on the left. Yet this classification would be overly simplistic, since there are many imperfections in the numeral-signs that defy a simple multiplicative explanation.

The solution to this taxonomic conundrum becomes entirely clear through analyzing the origin of the siyaq system. The siyaq numeral-signs are extremely reduced cursive versions of the corresponding Arabic lexical numerals. Because the Arabic lexical numerals are multiplicative-additive (just as in English and most other languages), when they were cursively reduced into abstract and non-phonetic siyaq signs, they retained a visual vestige of their original multiplicative nature. Thus, the Arabic word for 'thousand', *alf*, which is written phonetically as الف, is reduced to الـ in the siyaq numerals. Kazem-zadeh (1915) has demonstrated that the origin of practically all of the siyaq numerals can be understood in this manner. This unusual origin also explains the odd structural features of the siyaq numerals, such as the placing of the units before the tens in numeral-phrases. Yet siyaq numeral-phrases could not be read phonetically; they are all too reduced to be understandable except to those trained in the system's use. Especially in non-Arabic-speaking areas, the association between numeral-words and numeral-signs was quite limited. Thus, the siyaq system is clearly numerical notation, not a set of lexical numerals.

The earliest document containing siyaq numerals (in fact, the *dewani* variant used by the Arabs) is a list of expenses and receipts presented to the Abbasid caliph Al-Moktadir Billâh by his minister, Ali ibn 'Isa, dating to 306 A.H. (919 AD) (Kazem-zadeh 1915: 14). The numeral-signs from this text are already quite impossible to read as lexical numerals, so these signs may have been used even earlier. The numbers expressed in this text and later ones normally are used for representing monetary amounts, but in some cases were used to express weights as well as discrete quantities of objects (Kazem-zadeh 1915: 31-32). It appears that the primary reason why the siyaq numerals were chosen over the Arabic positional numerals or some other system was that in so doing, control could be exercised over who could read it. Kazem-zadeh (1915: 31) argues that the primary function of the siyaq numerals was to prevent financial corruption by making forgery more difficult. It is probable that the system served both functions, whatever its users may have intended.

From its origin as an administrative tool within the Abbasid caliphate, the siyaq numerals came to be used for many centuries in all of the major successor states to the

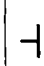
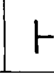
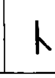
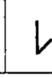
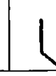
Abbasids. It appears to have been extremely popular in the Ottoman Empire and Safavid Persia between the 16th and the 18th centuries. The Ottomans apparently stopped using siyaq numerals some time in the 19th century; however, while Ifrah (1998: 543) reports that it ceased to be used *everywhere* in the 19th century, Kazem-zadeh (1915: 10) observed that it was still being taught in Persian schools at the turn of the twentieth century. It is not used today anywhere in the world and has not been for some decades.

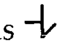
Cistercian

For most purposes, medieval European scribes used Roman numerals (in Western Europe) or Greek alphabetic numerals (in Eastern Europe), with Western numerals becoming increasingly frequent from the 11th century onward. Yet, beginning in the early 13th century, an unusual system began to be used on a limited number of manuscripts and marked on objects, primarily in contexts associated with the scribal tradition of the Cistercian monks. I therefore call this system 'Cistercian numerals', even though neither its earliest nor its latest users were Cistercians. While it has been known to antiquarians and historians for centuries, and has been studied continuously by palaeographers and historians of mathematics since the 1920s, this system has been ignored in all the synthetic works on numerical notation of the twentieth century. Thanks to the recent work of King (1995, 2001), which supersedes all earlier research completely, we now have more information about this system than we do for many that were used much more widely.

The precursor of the full-fledged Cistercian numerals was a set of eighteen symbols introduced by John of Basingstoke (John of Basing), archdeacon of Leicester, in the early 13th century (Greg 1924; King 2001: 51-57). These signs are shown in Table 10.15 (King 1995: 202).

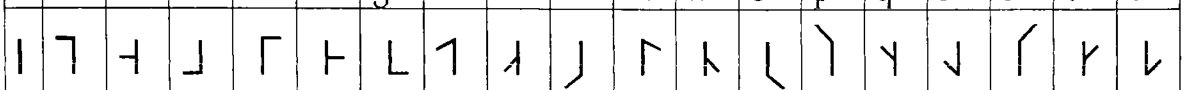
Table 10.15: Numerals of John of Basingstoke

	1	2	3	4	5	6	7	8	9
1s									
10s									

The symbols for the units can be grouped into three sets of three (1-3, 4-6, 7-9), based on the position of the short stroke to the left of the vertical stroke (at the top, middle, and bottom, respectively). Each of the tens signs is a horizontal mirror image of the corresponding unit-sign. This allowed the two exponents to be combined into a single sign, so that 75 would be written as . There is no way to write numbers higher than 99. There are two valid ways to classify this system. It may be considered a ciphered-additive decimal system, in which there are nine distinct signs for the ones and nine more for the tens. Alternately, recognizing that the signs for the tens are mirror images of those for the ones, we may consider this system a very peculiar ciphered-positional system - one in which the positions are not arranged in a simple line, but in which the *orientation* of the numeral-sign around the vertical stroke determines its value.

The origin of the Basingstoke numerals is still open to debate. One theory holds that they have their origin in Greece (King 2001: 57-65). Basingstoke's biographer, Matthew Paris, reported in his *Chronica maiora* that Basingstoke spent much time in Greece and learned the system from scholars in Athens. Moreover, a 4th century BC (!) tablet found on the Acropolis contains a form of cryptographic alphabetic shorthand whose signs are similar in shape to Basingstoke's numerals. Yet the ancient Greeks never used this shorthand to express numbers and there is no evidence for its survival in Byzantine scholarship. A second theory, one for which no great leap in time and space is required, is that a system of alphabetic shorthand known as the *ars notaria,* which developed and was used in England in the 12th century, inspired Basingstoke's invention (King 2001: 66-71). The *ars notaria* used all 18 of Basingstoke's numerals (plus a vertical stroke) to represent nineteen alphabetic signs. Moreover, while the *ars notaria* were not used to express numbers, when they are placed in alphabetic order and correlated with their numerical values in Basingstoke's system, a clear pattern emerges, as seen in Table 10.16 (King 2001: 68-69).

Table 10.16: Alphabetic and numerical values of the *ars notaria* / Basingstoke's system

a	b	c	d	e	f	g	h	i	l	m	n	o	p	q	r	s	t	u
																		
-	2	5	8	20	50	80	3	6	9	30	60	90	1	4	7	10	40	70

It is remotely possible that the graphic similarities and patterning of the *ars notaria* and Basingstoke's system were developed independently. Nevertheless, since we know the former to have been invented in the 12th century, the most parsimonious theory is that Basingstoke learned the *ars notaria* and then hit on the idea of assigning numerical values to these alphabetic signs. This raises the possibility, which King does not mention, that perhaps Basingstoke's travels in Greece taught him the Greek *alphabetic* numerals (ch. 5), leading him to hit on the idea of using *ars notaria* letters as numerals. Given that the alphabetic numerals are ciphered-additive, there is a structural similarity between the two systems. The use of alphabetic numerals was infrequent in Western Europe, but would have been common in early 13th century Athens. While it is impossible to confirm this theory, it does explain why Matthew Paris would have claimed them to have been inspired by the Greeks.

While Basingstoke's numerals appear in only two texts other than the *Chronica maiora*, one of these is a late 13th century manuscript from a Cistercian monastery, Whalley Abbey in Cheshire. This is relevant because the next place we find a system like Basingstoke's numerals is in late 13th century Cistercian manuscripts from France and Belgium. While these signs were slightly different from his numerals (and in fact differed considerably from manuscript to manuscript within the Cistercian tradition), they were clearly derived from the earlier English signs. The most common variant of this system is that shown in Table 10.17 (King 2001: 102).[9]

Table 10.17: Cistercian horizontal numeral-signs

	1	2	3	4	5	6	7	8	9
1s									
10s									
100s									
1000s									

Whereas Basingstoke's numerals had signs only for the units and tens, this more developed system included signs for the hundreds and thousands as well, and used a horizontal base stroke rather than a vertical one. Nevertheless, the structure of the system is essentially the same, only with four positions instead of two, with the units in the top left, the tens in the bottom left, the hundreds in the top right, and the thousands in the bottom right. Thus, even though it neither possessed nor needed a sign for zero, we may classify it as a positional system based on orientation, or as a ciphered-additive system for which signs for the same multiple of different exponents happened to resemble one another. Table 10.18 shows how several numbers would have been expressed using this system.

Table 10.18: Horizontal Cistercian numeral-phrases

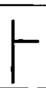
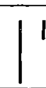
157	
2345	
6666	
9002	

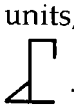
[9] King (2001: 39) provides a chart illustrating the enormous variation in this system within what was, after all, a very limited manuscript tradition.

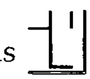
These numerals were used in a variety of Cistercian manuscripts from the 13th to the 16th centuries, primarily in the Low Countries and neighbouring regions of northern France (King 2001: 95-130). They were used extensively in the pagination of Cistercian religious texts and the numbering of sermons as well as for writing numbers (especially year-numbers) in the body of texts. These signs were much more compact than the corresponding Roman numerals in most cases, and, while they were in direct competition with the increasingly popular Western numerals, they seem to have been disseminated widely within the Cistercian scribal tradition. Given that these manuscripts were intended for a *very* limited audience and, since they sometimes included charts in the margin of the text explaining their use, there seems to be little possibility that they were used cryptographically at this period.

Starting almost at the same time as the horizontal numerals, a variant Cistercian system began to be used with vertical base-strokes; these signs were similar to the horizontal system, only rotated 90° clockwise, so that the units occupied the top right position. A common version of these signs is shown in Table 10.19, indicating only the units (King 2001: 39).

Table 10.19: Vertical Cistercian numerals

1	2	3	4	5	6	7	8	9
								
				(4+1)		(6+1)	(6+2)	(8+1)

The vertical signs for the tens, hundreds, and thousands are simply those for the units, flipped and rotated as in the standard system, so that 5107 would be written as . What is notable about this system is the use of an additive framework within each exponent to construct many signs, so that the ciphered signs for 5, 7, 8, and 9 are additive combinations of the other signs. This adds a level of transparency to this variant that is not present in other ciphered systems, including the Western numerals and the standard Cistercian numerals described above.

The vertical numerals first appear in a manuscript copied in Paris in the late 13th century, in which pages are foliated using this system (King 2001: 153-155). While they were not used as frequently as the horizontal numerals, they were employed in a wider set of contexts. They are inscribed on an astrolabe from Picardy, the only example we have for their use on an object rather than in a text (King 2001: 131-151). They are found in a 15th century arithmetical text from Normandy, where a technique is described for writing numbers higher than 10,000 by placing a sort of bracket around a lower number using the multiplicative principle, so that 126,000 would be written as  (King 2001: 159). They also occur outside of northern France on a late 15th century astronomical table from Segovia, which belongs to a set of eclipse computations by the Jewish Spaniard Abraham Zacuto (Chabás and Goldstein 1998). Perhaps most unusually, a few manuscripts from Bruges describe their use as markings on wine-barrels and wine-gauges, as part of the mercantile practices of vintners starting in the late 14th century and used as late as 1720 (King 2001: 164-171; 239-242). Unfortunately, no marked wine-barrels or related artifacts exist to complement the textual evidence for this mercantile practice. Thus, while the horizontal numerals were diffused only within restricted Cistercian circles, the vertical numerals were used in many non-Cistercian contexts, including scientific ones.

The advent of printing in the middle of the 15th century, and the decline in the Cistercians' fortunes that accompanied the Reformation, were disastrous for the survival of this system. The numerals ceased to be used regularly in the 16th century. Yet, just at this point, interest in the numerals from an academic and mystical perspective began to arise. They appear as historical curiosities in many 16th century texts, most notably *De occulta philosophia* (1531-33) by Agrippa of Nettesheim (King 2001: 190-202). They are also found in *De numeris* (1539) by Johannes Noviomagus and *De subtilitate libri XXI* (1550) by

Girolamo Cardano, both of which cite Agrippa as an authority.[10] In these texts, and frequently thereafter, the numerals were mistakenly thought to be "Chaldean", an appellation often used in the Renaissance to refer to mystical learning supposedly diffused from the Near East, especially Babylonia. Even Cajori (1928: 68-69) cites the Chaldean theory of Agrippa, though he is rightly dubious that this tells us much about their true origin.

The use of the Cistercian numerical notation system in well-known mystical and mathematical texts ensured that they were never completely forgotten, even though knowledge of their true origin was lost. They were described in various works on magic, the occult, and astrology, as well as in a variety of early works on numerical notation (King 2001: 210-238). Yet, other than wine gauging in Bruges mentioned above, they seem to have been *used* only rarely after 1550. A possible exception is that a group of Parisian Freemasons seem to have used the numerals in some of their private correspondence with fellow members in the 1780s (King 2001: 243-246). The last non-scholarly mention of the numerals was by a number of German nationalistic authors in the early twentieth century, who saw the Cistercian numerals as a sort of proto-Aryan runic numeration (King 2001: 251-261).

<u>Ottoman cryptographic</u>

Throughout the period of Ottoman dominance in the Middle East, between roughly 1450 and 1900, the standard Arabic positional numerals (ch. 6) were by far the most common system in use, while various ciphered-additive systems were used in certain contexts, most notably the Arabic abjad numerals (ch. 5). In addition, a number of quasi-cryptographic systems that bear little to no resemblance to the Arabic alphabetic or

---

[10] Curiously, however, Noviomagus lists the horizontal rather than the vertical numerals, and some of Cardano's vertical numeral-signs are more similar to Basingstoke's 13th-century numerals than they are to the later vertical signs.

positional systems were used by Ottoman administrators (particularly military clerks). Four of these were reported in Western scholarly literature in 1899 by M.J.A. Decourdemanche; the extent of my knowledge of them is based entirely on his research. While the historical importance of these systems was not great, their importance lies in their structural ingenuity: three of the systems described by Decourdemanche have unusual structural properties. A fourth, which I will not describe, is no more than a set of cryptic alphabetic signs different from those used in the abjad numerals.

The first of these systems, known as *keutuklu*, was used by clerks to record data concerning the recruitment of Christian youth into the Ottoman army (Decourdemanche 1899: 261). The signs of this system are shown in Table 10.20 (Decourdemanche 1899: 260).

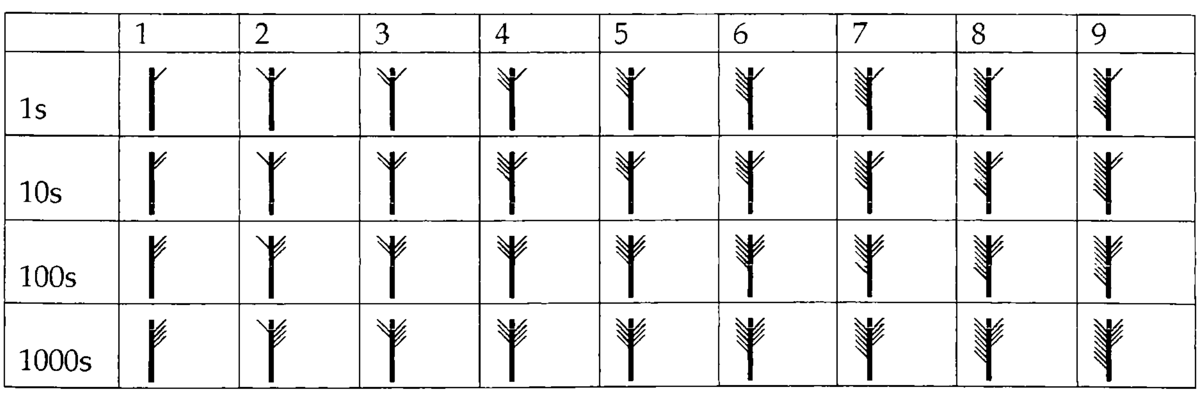
Table 10.20: *Keutuklu* numerals

	1	2	3	4	5	6	7	8	9
1s	ǀ	┘	┐	┌	└	┤	┬	├	┴
10s									
100s									
1000s									

This system has the appearance of a decimal ciphered-additive system, given that there are unique signs for each multiple of each exponent of 10, and this is certainly a valid interpretation. The numeral-signs themselves are not arbitrary; instead, they are constructed by adding small circles to the set of nine basic unit-signs (one circle for the tens, two for the hundreds, and three for the thousands). An alternate way of looking at this system, then, would be to regard it as *multiplicative*-additive, with the unit-signs being the basic linear frames for 1 through 9 and the exponent-signs being one, two, or three circles. Additionally, there are graphic resemblances among the signs for 2 through 5 (which are based on the sign ┘ and transpositions thereof) and among the signs for 6 through 9 (based on), although these do not affect the system's structure. However one chooses to look at it, there is a clear graphic resemblance between the signs in each column (e.g. 6, 60, 600, and 6000), which dispels one of the objections often levelled at ciphered-additive systems, that being the arduousness of memorizing many different signs. Because Decourdemanche does not describe how these signs combined with one another to produce numeral-phrases, we do not know whether 4269 would have been written from left to right as in the Arabic positional numerals ( ), from right to left as in the Arabic abjad and other alphabetic systems ( ), or in some other manner.

The second system I will consider, known as *ordoui* ("army"), is also quite ingenious in the way numeral-signs were constructed. The most common variety of this system, known as *ordouï cheïlu* "army equipment", was used by the Ottoman army for enumerating provisions, equipment, and other military supplies (Decourdemanche 1899: 262). The signs of the system are shown in Table 10.21 (Decourdemanche 1899: 263).

Table 10.21: ***Ordouï cheïlu*** **numerals**



Like the *keutuklu* system, the *ordouï cheïlu* can also be classified as ciphered-additive or multiplicative-additive, and it is clearly a decimal system. Each sign consists of a vertical stroke with a number of diagonal strokes leading off it to the left and right. The left side represents the number of units, with no strokes for 1 up to eight strokes for 9, while the right side indicates the exponent, with one stroke for the units, two for the tens, three for the hundreds, and four for the thousands. While this system seems to have a cumulative component at first glance, the strokes on the left do not add up directly to the number of units, but rather one fewer than the number of units represented (zero for 1, one for 2, ... eight for 9).[11] Again like the *keutuklu*, it is possible to derive the value of a sign easily from a limited set of basic rules. In addition, it is possible to see how this system might be rendered potentially infinite, even though it is non-positional, using five strokes on the right side for the ten thousands, six strokes for hundred thousands, and so on. Decourdemanche does not report how the signs of this system were arranged into numeral-phrases.

A variant of the *ordoui* system was used for recording the numerical strength of military units, and could additionally serve as a cryptographic script. The signs of this variant are shown in Table 10.22 (Decourdemanche 1899: 262).

Table 10.22: *Ordoui* numerals for personnel

	1	2	3	4	5	6	7	8	9
1s									
10s									
100s									
1000s									

The signs of this system are highly irregular in comparison in comparison with the *ordouï cheïlu*. Instead of indicating the exponent of the sign by the number of strokes on the right, the signs are grouped erratically in sets of three or four (1-4: one right stroke; 5-7: two right strokes; 8-10: three right strokes ... 800 – 1000: eight right strokes). Moreover, there is no common feature among the multiples of different exponents, so

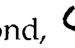
[11] While the *total* number of diagonals (left and right) equals the relevant number in the ones column, this pattern does not hold for the higher exponents.

that 4, 40, and 400 have no inherent similarity. Finally, whereas the *ordouï cheïlu* could be used to write any number up to 10,000, the highest sign in this system was 1,000. Yet the nature of this system becomes clear when its signs are correlated with the Arabic abjad (ch. 5). The 28-sign abjad was divided into eight mnemonic groups of three or four signs apiece, and the numerical values assigned to the abjad correlate perfectly with the divisions of this system. Moreover, the numeral-signs above could be used not only in their numerical sense, but also to represent the appropriate letter of the Arabic abjad. This function would not have been easy to reconcile with the *ordouï cheïlu*, which was structured strictly according to exponent, and had 36 rather than 28 signs.

The third notable Ottoman cryptographic system, known as *damgalu* "inspection", was used for marking numerals on military equipment, and also could be used as a cryptographic script (Decourdemanche 1899: 264-265). The signs of this system are shown in Table 10.23 (Decourdemanche 1899: 265).

Table 10.23: ***Damgalu*** **numerals**

	1	2	3	4	5	6	7	8	9
1s									
10s									
100s									
1000s									

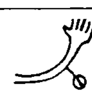
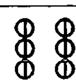
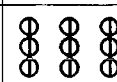
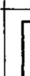
Unlike the *keutuklu* and *ordoui* systems, the *damgalu* numerals have only twenty-eight signs, like the Arabic abjad numerals, corresponding to 1-9, 10-90, 100-900, and 1000. Instead of simply using alphabetic signs, however, each sign of this system has four registers, in each of which a line or a dot is placed. This allows for 16 (2^4) combinations, which means that some other technique was needed to represent the last 12 signs. This was done by using two additional signs. The first, , was placed under any sign whose bottom register was occupied by a line, while the second, , was placed under signs whose bottom register was a dot. The signs shown in Table 10.23 are the variety of *damgalu* used by the Ottoman navy, while a separate system was used in the army, which was identical except that it used the 16 combinations in a different order. There is no correlation between the sequence of dots and lines and the numerical values in question, so the *damgalu* is particularly cryptographic. It cannot be considered multiplicative in any way, and is a simple decimal, ciphered-additive system. The use of additional signs for the last 12 numerical values is not particularly significant from a structural perspective; it was simply a necessity imposed by the lack of adequate signs available with the 16 basic combinations. Like the *ordoui* variant for personnel, the *damgalu* numerals were each correlated with one of the 28 signs of the Arabic abjad, and could be used to stand for phonetic values as well as numerical ones.

We know remarkably little about these systems' origin or history, save that they were employed in the nineteenth century, when Decourdemanche reported on their use. They are probably related to the ciphered-additive alphabetic numerals of the region, of which the Arabic abjad numerals were the most common. This ancestry is almost certain for the *ordoui* and *damgalu* systems, which were organized according to the structure of the Arabic abjad and could stand either for numerals or for the corresponding letter of the script. In the case of the *keutuklu* system, it is possible (though unproven) that the Arabic positional numerals were partly responsible for its unusual structure. Decourdemanche reports that at least some of them were used (at least by some individuals) in the nineteenth century, but how old they are is quite unclear. They are treated by Decourdemanche as already being obsolete at the turn of the 20th century, and do not appear to have survived past the end of the Ottoman Empire.

Easter Island

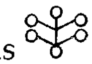
Among the many mysteries concerning the undeciphered scripts of Easter Island, of which the most common is the classical script or *rongorongo*, is whether there was any numerical notation associated with it. The rongorongo script was used for writing the Rapanui language of the Polynesian inhabitants of the island; it may be a syllabary with a slight logographic component (Macri 1996). Yet decipherments of the Easter Island scripts are at best incomplete, and they rival those given for the Indus and Minoan Linear A scripts in their use of conjecture.[12] It is possible that rongorongo signs for various marine mammals symbolized particular numbers from 1 through 9, but this interpretation, if correct, tells us more about Rapanui numerology than about numerical notation (Barthel 1962; Schuhmacher 1974). Elsewhere, Barthel (1971: 1175) explicitly denies that there are rongorongo numeral-signs. The only plausible theory that has been raised concerning the possibility of Easter Island numerical notation is that presented by Bianco (1990). Some of the signs that he has suggested might be part of a rongorongo numerical notation system are shown in Table 10.24 (Bianco 1990: 41).

Table 10.24: Putative Easter Island numerals

1	2	3	4	5	6	7	8	9	10
									

The signs shown represent only a fraction of the variation that Bianco believes may have existed in the numerals; for instance, he lists 12 different possible signs for '1'. This putative system obviously has a cumulative component in the use of multiple signs (often circles with vertical lines through them as shown, but also sometimes diamonds, and also sometimes without lines through the signs). It is decimal in that it has a sign for

[12] See Fischer 1997 for a remarkably complete summary of dozens of decipherment attempts from the 1860s to the present.

10, and at least minimally quinary in the use of the hand (with a circle and line) for 5, though there are no cases where the sign for 5 combines with other numbers of units. Bianco (1990: 40) suggests that in some cases, alternate signs were used, such as  for 6. No signs can clearly be identified as representing 7 and 8. In a single instance, a sign  is attested that, according to Bianco's (1990: 46) interpretation, represents (3x10)+5, or 35. If so, this system would be multiplicative-additive and decimal, with a quinary component for 5.

Nevertheless, I am rather unconvinced by this interpretation, for several reasons. First, if this truly comprised a numerical notation system, we would expect to find combinations of two or more numeral-signs, and of various numeral-signs with the same non-numerical signs, far more frequently than is attested from the texts. In fact, we do not find combinations of signs with any great frequency, and the single signs do not match up regularly with other non-numerical signs (such as those that might be logograms for objects being counted). Second, the assumption that a grapheme that consists of a group of identical signs (in this case, circles) is highly likely to represent a cumulative numeral-sign is quite dubious, especially because there are no instances of 7 or 8 being expressed as simple cumulative signs. I am particularly dubious of interpretations where the signs are not simple concatenations of identical signs, but, as in the alternate sign for 6, are constructed by using various joining lines. Third, the use of the hand as a sign for 5 only makes sense if it combines with the unit-signs for 1 through 4 to represent 6 through 9, which it never does. Fourth, the relative frequencies of these signs in the tablets do not give me great confidence that they are numerical. There are 150 examples of the twelve different signs for 1, 70 examples of 2, and 430 of 3, but none for 7 or 8, and only three for 9. Finally, the very promising analysis of the rongorongo script by Macri (1996) assigns grammatical functions to many of these signs, making it unlikely that they also served as numerals. Similarly, Fischer's (1997) recent decipherment attempt concurs that the rongorongo script had no identifiable numerical notation.

While no one of these factors is devastating to Bianco's case, in combination, they lead to the conclusion that we have no idea whether the rongorongo script had a corresponding numerical notation system. Unlike other undeciphered scripts, such as Linear A, for which the numeral-signs are obvious and frequent, rongorongo texts lack any of the markers that would help identify such signs. Despite Bianco's (1990: 39) statement that "[i]l est normal de trouver un système représentatif des nombres dans une écriture ancienne, les tablettes pascuanes ne pouvaient échapper à cette règle générale", there exists no iron law that every script must have its own numerals. It is quite possible that the texts were not used for functions in which numerical notation was necessary or useful. Alternately, it may have had a ciphered rather than a cumulative system, in which case we would be unable to identify numeral-signs until the script is deciphered more fully. A final possibility is that so few rongorongo texts survive (about two dozen) that any system that may once have been recorded is now lost. Until further evidence becomes known, the numeral-signs of Easter Island will remain a mystery.

Indus

The writing system of the Harappan civilization, centered in the Indus River valley is one of the great remaining mysteries in the field of script decipherment. It was used from around 2500 BC to 1900 BC on several thousand very short inscriptions (averaging five signs per 'text'), and was written primarily from left to right (Parpola 1996). Unfortunately, there is no reliable basis on which to decipher the script, given that the language it represents is not known (though it is often supposed to have been a Dravidian language), and there are no bilingual inscriptions. The situation is even more grave than that for deciphering scripts such as Linear A, for which we have many easily readable numeral-phrases and associated ideograms (see ch. 2). As a result, the Indus script has been subjected to a variety of ridiculous interpretations.[13] We have barely enough evidence to confirm the existence of a numerical notation system in the ancient Indus Valley, much less determine its origin, history, or function.

There have been several earnest attempts to decipher the Indus numerals, mostly relying on the very frequent occurrence of groupings of vertical strokes on the inscriptions. Table 10.25 shows these numerals as well as the frequency with which they are encountered in the texts (Fairservis 1992: 62).[14]

Table 10.25: Short and long Indus strokes and frequencies

	1	2	3	4	5	6	7	8	9	10
Short strokes	I	II	III	IIII	III II	III III	IIII III	IIII IIII	IIIII IIII	IIIII IIIII
	?	?	151	70	38	38	70	7	2	1
Long strokes	I	II	III	IIII	IIIII	IIIIII	IIIIIII			
	149	365	314	64	22	3	6			

It is generally agreed that in many cases, these signs represent low numbers in a cumulative fashion; the short strokes are grouped into sets of three, four, or five, just as the signs of most other cumulative systems. The longer ungrouped vertical strokes only occur in the early Indus inscriptions; during its mature phase, the shorter strokes are used exclusively (Parpola 1994: 82). Because these sets of strokes are paired interchangeably with non-numerical graphemes (e.g. the 'fish' sign  is attested in combination with 3, 4, 6, and 7 strokes), we can be relatively confident that they could be

[13] Most notable among these, from the perspective of numerical notation, is Subbarayappa's (1996) claim that every one of the 200 or more Indus signs has a distinct numerical value, using cumulation, ciphering and multiplication haphazardly, and invoking parallels with practically every civilization of South and East Asia as well as with Mesopotamia and Greece to justify his unlikely theory.

[14] Fairservis (1992: 183) provides no count of single and double short strokes because these are also assigned grammatical functions (as genitive and locative case markers, respectively) in his decipherment.

used as numeral-signs (Parpola 1994: 81). Yet Ross (1938) long ago pointed out that some groupings of vertical strokes pair non-interchangeably with other signs, which suggests that they may have had phonetic or grammatical values in these instances (Ross 1938; Fairservis 1992: 12). This is parallel to the frequent use of numeral-signs phonetically in Chinese writing, and resembles abbreviations such as "K-9" for *canine* in English. Thus, we must be very cautious before attributing a numerical function to all these signs. This is particularly true in the case of the Indus symbol , which occurs many times but never in the same contexts as other numerals do; Fairservis (1992: 71) argues that it should be read as 'rain', which may or may not be correct, but is far more likely than '12'. The Indus texts are so short and so devoid of contextual information that we ought to be very careful not to read too much numerical information into them.

One of the problems with this interpretive framework for the Indus numerals is that it does little to establish whether this system had a base and used an intraexponential principle to write higher numbers. Fairservis notes that there is a sharp drop-off in frequency after 7 for both the long and short vertical strokes, and that in fact there are no attested instances of 8 or more long strokes. From this, he concludes that the Indus numerals were probably octal or base-8 (Fairservis 1992: 61-2). Perplexingly, however, he then proceeds to assert that there are pictographic signs for 8, 9, 10, and 11 (, , , and , respectively) that were simultaneously numerical and calendrical, indicating the eighth through eleventh months of the (as-yet unknown) Harappan calendar, because these four signs, along with vertical strokes for 1 through 7, are found in association with a sign that he thinks represents 'month' (Fairservis 1992: 65). This theory has not been widely adopted by scholars of the Indus script (cf. Pettersson 1999: 103), and since it provides no evidence of signs for a base or exponents of that base, does nothing to prove the existence of octal numerals. While I accept that vertical strokes probably indicated low integers in the Indus script, this fact does not tell us anything about the structure of its numerical notation system.

Our best evidence for a legitimate Indus numerical notation system is not found on the seal inscriptions, but rather on nine inscribed potsherds and copper and bronze tools found at Mohenjo-daro, Canhujo-daro, and Kalibangan. Many of these objects are inscribed with sets of vertical strokes (|), crescent shapes (∩), and sometimes other script-signs. These two signs are sometimes found in combination on the seal inscriptions, but never in large numbers and never so clearly separated from the rest of the text. Pettersson (1999) adds that, in addition to vertical strokes and crescents, a distinction needs to be made between vertically and horizontally oriented strokes. One object, a chisel or axe blade (DK-7535[15]) from Mohenjo-daro, contains all three signs, as shown in Figure 10.6 (Parpola 1994: 108).

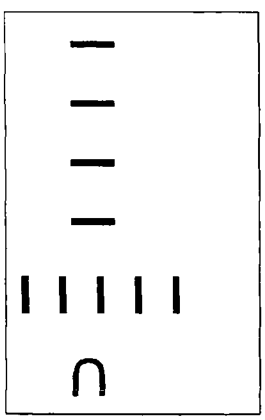


Figure 10.6: Inscription on artifact DK-7535 from Mohenjo-daro

While I am reasonably convinced that inscription and similar ones on other Harappan tools are numerical in function, there is no agreement as to the specific structure and value of the signs. Fairservis (1992: 67-69) has constructed a convoluted argument whereby the vertical strokes (standing for units) can serve either an additive or multiplicative role in the numeral-phrase depending on whether they follow or precede the crescent sign(s). Pettersson (1999: 102-103) points out that there are no cases where a crescent sign is both preceded by and followed by vertical strokes, thus making this

[15] There is some confusion over the identification of this object, which is assigned different artifact numbers by Parpola (1994) and Pettersson (1999).

theory unlikely. Fairservis (1992) and Pettersson (1999) argue that because none of these nine objects contains more than seven of any sign, the Indus numerals must be octal rather than decimal. Yet Parpola (1994: 82) argues that the crescents probably represent 10 rather than 8. Either of these interpretations of the system would mean that the Indus numerical notation system was cumulative-additive.

At present, I do not think there is enough evidence to decide whether the crescent-sign had a value of '8' or '10'. Nine numeral-phrases is a very limited corpus from which to conclude that, since no sign is repeated more than seven times, the numerical base must be 8. Fairservis' sign count from the seal inscriptions (see Table 10.25 above) is possible evidence in favour of the octal interpretation, but the lack of numeral-phrases with different signs combined additively in this genre of texts does not give me great faith that there was a unique sign for 8, much less for 64, 512, and so on. At the same time, there is insufficient support for the decimal hypothesis, other than a universal presumption in favour of decimal interpretations of numerical notation systems. The very limited linguistic reconstructions regarding Proto-Dravidian (presuming that the Harappan language was a member of the Dravidian family) are very ambiguous and tenuous, but seem on balance to support a decimal interpretation, since roots for 'ten' and 'hundred' have been reconstructed (Parpola 1994: 169). The linear measures of the Harappans appear to have been decimal (Sarton 1936), and the system of weights is partly decimal and partly binary (Parpola 1994: 169; Pettersson 1999: 106). None of these objects bearing numeral-signs has any inscriptions on it, numerical or otherwise (Pettersson 1999: 91). Pettersson's (1999) attempt to correlate the numerical signs on the metal tools with their weights showed only that no metrological interpretation of their meaning (either decimal or octal) was likely to be correct.

The origin of these numerals is unknown at present. They are entirely unlike the Sumerian numerals (ch. 7) used in Mesopotamia at the time of the invention of the Indus script. It is interesting that the Egyptian hieroglyphic numeral-signs for 1 and 10 are |

and ∩, respectively, but, even if the Indus crescent-sign represented 10, this similarity could have arisen by chance. While there are vague similarities in the metrological systems of Egypt and the Indus Valley, I know of no evidence of cultural contact between the two regions (Petruso 1981). It is probably best to assume at present that the Indus numerals were independently invented.

The fact that the Indus script is completely undeciphered, coupled with the limited number of surviving numeral-phrases, makes it nearly impossible to identify the function(s) for which they were used. There is certainly no evidence of the numerals' use for accounting or administration, which is abundant for other undeciphered scripts, such as Linear A and Proto-Elamite. The wide variety of materials on which numerals are found (clay seals, potsherds, metal tools) suggests that it was used widely among literate Harappans, but even this hypothesis requires caution. The Harappan civilization declined precipitously after 1900 BC, although it may have survived in certain regions for a century or two longer. Despite claims to the contrary, there is no evidence that the Indus numerals had any influence on the Brāhmī numerals (ch. 6), which arose almost 1500 years later.

<u>Summary</u>

Because the systems described in this chapter are not part of a single phylogeny, they share little in common, save that they do not form part of any larger family of systems. It is worthwhile to note, however, that few of these systems have been studied in any histories of numeration. While systems such as the Inka quipu have been ignored or belittled because they are not associated with a script, a formal analysis of their properties shows them to be numerical notation systems that are comparable to any other. The failure to recognize other systems, such as the Cherokee, Pahawh Hmong, and various African systems, is probably a result of the marginalization of these societies in the modern world-system. If they are recognized at all, such systems are thought to be unimportant because they are derived from Western numerals and because they often die out rapidly. This is unfortunate, because these systems are very often structurally distinct from Western numerals, contradicting the assumption that African and other cultures influenced by Western imperialism are essentially devoid of independent scientific and numerical achievements. Moreover, they are important if we want to understand the sorts of circumstances that lead to the development of new systems. In most of these cases, Western, Arabic, or other positional systems were available to be adopted. The fact that indigenous systems were invented suggests that the desire to resist imperialistic institutions or to produce local alternatives to foreign inventions may be motivating factors governing the development of these and other systems. The fact that several numerical notation systems have been invented in the twentieth century, without even considering various systems such as hexadecimal numeration used in computing (a topic beyond the scope of this thesis), suggests that the creation of numerical notation systems remains an ongoing project.

Chapter 11: Cognitive and Structural Analysis

In Chapters 2 through 10, I described over 100 different numerical notation systems spanning over 5000 years and every inhabited continent. The raw data on these systems are summarized in Appendix A. I have shown that, while there are historically determined similarities among the systems of each phylogeny, numerical notation is a phenomenon in which the same structures and principles emerge independently multiple times. This situation creates a paradox only if we cling to the dichotomous assumption that historical explanations must be completely particularistic and must stand in stark contrast to universalizing ones. Armed with the data amassed throughout this study, I am now prepared to demonstrate how a combination of particularist and universalist approaches permits the explanation of the synchronic and diachronic historical patterns I have documented. In this chapter, I will elucidate numerous regularities (a large number of which are universals) and then suggest a number of cognitive factors that help to explain why systems are the way they are, and why they change in the ways that they do. Yet the analysis of the structure of numerical notation systems, while important, is insufficient as a full explanation of these patterns (particularly evolutionary patterns of change), because social context and historically contingent events have played an important role in various episodes in the history of numerical notation. In Chapter 12, I will look at social and functional explanations for certain patterns of change in systems, and then combine socio-functional and cognitive-structural factors into a single explanatory framework for the attested historical patterns I have found.

There are some domains of human experience for which the role of contingency is so great, or the functional constraints so minimal, that we cannot speak meaningfully of regularities or laws. Numerical notation is not one of them. Using the data from this study, I have discovered approximately 30 regularities that apply to numerical notation systems.[1] These regularities concern which kinds of systems are attested and which are not. These can be subdivided in various ways, but the most important division, for my purposes, is that between synchronic regularities, which apply to numerical notation systems considered as unchanging entities, and diachronic regularities, which apply to relations between systems over time. Because the unit of analysis is different for diachronic and synchronic regularities, the sorts of explanations they require will necessarily be different. Nevertheless, both synchronic and diachronic regularities can be universals (for which there are no exceptions), or statistical regularities (which hold true only for a preponderance of cases). While true universals are usually more notable than statistical regularities, we need not deny that statistical regularities are important, and may in fact be caused by cognitive factors similar to those that produce universals. Whenever there are exceptions, I have as an expositionary device stated the regularity in its universal (exceptionless) form and then discussed the exceptions in the text. It is useful to examine these exceptions to see whether they are trivial or, more enticingly, if they help explain why the rule exists in the first place. It is important to distinguish statistical regularities (general patterns that have exceptions) from implicational regularities, which take the form "If system A exhibits feature X, it will also exhibit feature Y" that apply to only a sub-section of the universe of numerical notation systems. Implicational regularities can be exceptionless or can have exceptions. Frequently, systems to which an implicational regularity does not apply do not actually *violate* it; rather, the feature of the system in question does not exist in outlying systems.

As discussed in Chapter 1, the systems that have been included in this study are those that are attested in the ethnographic or historical record. The clever skeptic can imagine systems that violate any of the regularities below, and some such systems have already been invented by scholars (Dwornik 1980-81), cryptographers (Wrixon 1989: 103), and science fiction writers (Pohl 1966: 179-192). This does not demonstrate that these

[1] See Appendix B for an unannotated list of the regularities described in this chapter.

generalizations are not 'true' regularities, or even that they are not universals, but merely proves that they are not logical necessities. Because the systems I studied satisfy these rules, even though it is not logically required that they do so, we must look instead to psychological and utilitarian constraints as the source of both the universal and the statistical regularities. That these constraints are apparently so great as to produce absolute universals among 100 or more structurally distinct numerical notation systems confirms the power of the mind (working in conjunction with the perceived environment) to constrain the structure of numerical notation systems. These correlations cannot be ignored, and because they deal with structural features of systems, they must be explained (at least partially) with reference to those features.

These regularities take on an even greater significance when they are compared to the set of regularities that apply to lexical number words. Greenberg's research has been of particular use to me in formulating the list of regularities below (Greenberg 1978). Where appropriate, I have indicated the correlations between my regularities and those he found for lexical numerals, without confirming or denying the validity of the latter set. However, the number of regularities for lexical numerals that do not apply to numerical notation systems, and vice versa, is quite striking. For every instance in which there is a parallel between lexical numerals and numerical notation, there is another in which there are significant differences between the two domains. Because of these differences, the regularities of numerical notation systems cannot possibly be derived from a biologically hard-wired 'universal grammar'. Since I have not conducted a comprehensive study of the lexical numeral systems associated with various numerical notation systems, any conclusions I have drawn relating to the connection between the two forms of representation must be regarded as tentative.

Synchronic Regularities

Synchronic regularities describe features that are common to all systems (All systems are X), which I call general regularities, or ones that are common to all systems of a given structure (If system A exhibits X, it will also exhibit Y), which I call implicational regularities. In either case, the numerical notation system is the basic unit of analysis. Either general or implicational regularities can have exceptions, although I have not included regularities that have many non-trivial exceptions that may suggest that the 'rule' is simply a coincidence.[2] I will begin with a brief list of axioms, which frame the phenomenon of numerical notation according to the basic guidelines set out in Chapter 1, before describing the general and implicational regularities I have been able to discover. I then list a small number of non-universals, which are statistical regularities whose exceptions are more interesting theoretically than are the systems that obey them.

Axioms

A1. All numerical notation systems can represent natural numbers.

A2. All numerical notation systems have a base.

A3. All numerical notation systems use visual and primarily non-phonetic representation.

A4. All numerical notation systems are structured both intraexponentially and interexponentially.

These features have been described fully in Chapter 1, and require no particular attention here, except insofar as they form the basis from which all other regularities are derived. Any representational system that does not conform to these four rules is not a numerical notation system, by my definition. Like the other regularities I discuss,

---

[2] Because there are too few numerical notation systems (and certainly too few independently invented systems) to warrant the use of most statistical techniques, I have not employed them. Judgements concerning whether a statistical regularity is significant or not are better made by considering the nature of the exceptions to it.

however, they are not logically necessary; for instance, it is possible to have a system that has no base.

General Regularities

G1. Any system that can represent N+1 can also represent N, where N is a natural number.

This is a universal, which I call the *Continuity Principle.*[3] It establishes the continuity of the sequence of natural numbers starting at 1, but does not imply that all numerical notation systems are infinite in scope. It also leaves open the question of the expression of zero, negative numbers, and fractions. It is conceivable that a system might be developed for the sole purpose of recording a set of non-sequential numbers with religious significance, or that a group of users would use one system for representing odd numbers and an entirely different one for even numbers. Such unusual systems have never been implemented. I suggest that one of the crucial representational functions of numerals is enumerating things in an ordinal sequence, for which only a continuous set of integers will suffice. This rule is so important that it might be argued that it should be made a definition of a numerical notation system, but I choose to leave open the possibility that a numerical notation system might not express a continuous set of natural numbers.

G2. All systems have a base of 10 or a multiple of 10.

This is a universal, which I call the *Rule of Ten.* It is possible that the Indus Valley civilization had an octal (base-8) numerical notation system (ch. 10), but as I discussed, the base of this system is not certain. Systems for representing fractions, which often use a different base than the systems for integers with which they are used, often have non-decimal bases, such as the base-2 Egyptian "Horus-eye" fractions (ch. 2) and the base-24

[3] See also Greenberg (1978: 254-255) for a similar principle concerning lexical numerals, which he calls the "thesis of continuity".

Linear A fractions (ch. 2). The only widely-used potential exceptions to the Rule of Ten are the binary, octal, and hexadecimal systems used in computing, but these show no signs of achieving wider currency as the general system of any society. The fact that some systems have sub-bases or extraneous structuring signs that are not multiples of 10 is irrelevant to the validity of this principle. The explanation of this feature requires that we consider several hypotheses, as well as incorporate additional regularities into the analysis. I will consider this question below (see 'Fingers and Numbers').

G3. All systems form numeral-phrases through addition.

G4. No system forms numeral-phrases through division.

These two rules are universals. Addition will always be found among the arithmetical steps by which a system is used to derive the values of numeral-phrases, whether it is the only operation (as in cumulative-additive and ciphered-additive systems) or not (multiplicative-additive, cumulative-positional, and ciphered-positional systems). It is possible to imagine a system that is purely multiplicative - for instance, one that expresses all numbers as prime numbers or the product of prime numbers - but this has never occurred. This rule does not imply, however, that every numeral-*phrase* in a system uses addition; the units and the exponents of the base are expressed with single signs in many systems, and thus do not involve addition. Addition is frequently combined with multiplication, which as a form of repeated addition is a very effective means of expressing large numbers, whether the exponent multiplier is explicit, as in multiplicative-additive systems, or implicit, as in positional systems.

While addition and multiplication are quite common cross-culturally, subtraction is extremely rare (being found only in the Roman numerals and a few Mesopotamian systems), and division is absent entirely from the operations used to form numeral-phrases for integers. It is certainly possible to imagine 50 and 10 being expressed as "2 100" or "2 20", but this is not attested. Lexical numerals only use division in the form of multiplication by ½ or ¼, and this is very rare (Greenberg 1978: 261). Even this operation is never found in numerical notation. There is of course the *physical* division of Etruscan, Roman, and Ryukyu tallying-based signs (e.g. Roman V (5) is the top half of X (10)), but this is a non-arithmetical technique governing the formation of signs. In any event, this graphic technique can be interpreted as doubling or halving with equal validity. Of course, this regularity does not deny that fractions can be expressed in numerical notation. The absence of division (and the rarity of subtraction) may be a result of the way humans think about number, a matter of representational convenience (since the use of divisive numeral-phrases would involve using large divisors to express smaller numbers), or a consequence of the rarity of such operations in lexical numeral systems.

G5. All numerical notation systems are ordered and read from the highest to the lowest exponent of the base.

This is a near-universal, which I call the *Ordering Principle*. Positional systems could be read from the lowest exponent to the highest, but this never occurs. While the exponents of purely additive systems could be placed in any order (e.g. in classical Roman numerals, which do not normally use subtraction, 217 could be written as IICVCX), this is never the rule, and occurs only when the writer has made an error. Subtractive forms such as the modern Roman numeral IX for 9 do not violate this principle, because they involve intraexponential structuring only. A fair number of lexical numeral systems do not always obey this principle, including many of the major European languages (e.g. Italian *sedici* = 6+10 although *diciasette* = 10+7). However, numerical notation systems almost never do so, instead reserving "low-high" forms for subtractive or multiplicative purposes. Users of a system immediately know, upon encountering a lower numeral-sign followed by a higher one, that an operation other than addition is involved. The Ordering Principle also applies to sub-bases and other forms of intraexponential addition, so that sub-bases always precede signs for the next lower full exponent. There are a couple of minor exceptions to the Ordering Principle; in certain systems of the Alphabetic family, including the Greek, Glagolitic, and Cyrillic alphabetic numerals, the numbers 11-19 are often written with the sign for 10 at the end of the numeral-phrase (e.g. Cyrillic 12 = ВІ (2+10), not ІВ (10+2)). These exceptions reflect the word order of the lexical numerals of the languages of these systems' users.

G6. No system uses signs for the operations used to derive the value of a numeral-phrase.

This is a universal. Even though all systems form numeral-phrases through addition, and many of them also use multiplication, this is always implied, never directly represented with a sign. This is in contrast to lexical numerals, in which it is very common to express at least some operations with words such as "sechshundert-fünf-*und*-vierzig", "duo*de*viginti", and other phrases that are even more complex. In fact, lexical numerals almost never express subtraction *without* some indication of the operation (Greenberg 1978: 258-259). Such signs in numerical notation would render numeral-phrases less concise without providing any additional clarity as to the phrase's meaning, which is already determined by the system's principle. The (near-) universality of the Ordering Principle means that the operation to be used can be inferred easily from the context. When lexical numerals show arithmetical operations explicitly, it is often *because* unusual ordering is being employed, as in the two numerals above.

G7. The only visual features used to determine the numerical value of figures in numerical notation systems are shape, quantity, and position.

This is universal or perhaps nearly-universal. It states, in other words, that the relevant features for determining the value of a numeral-phrase are the shapes of the particular signs used, the quantity of those signs, and their position. The colour of the signs, their relative size, and other extraneous graphic features do not affect the value of the phrase. It is certainly possible to conceive of an additive system where different registers of sign sizes, rather than their position, would determine the value of signs within numeral-phrases, thereby eliminating the need for a zero-sign. We could, for instance, use Western signs from 1 through 9 to write 462 as 462 and 402 as 42. A similar system could apply different colours to different exponents of the base. An exception to this rule is that in the proto-cuneiform and archaic Sumerian systems (ch. 7) the sign for 60 is a large version of the sign for 1. A second partial exception is the use of red and black-coloured rods in the Chinese rod-numerals (ch. 8), but this was only done occasionally and only served to distinguish positive from negative numbers. Certainly, no more than three features are needed for any system to distinguish any number from any other. Especially desirable features might be those whose different values are easily differentiated visually and those that are easily represented in writing. Size and colour do not bear numerical values because using these features would be extremely difficult to use – for instance, requiring users to employ many different-coloured inks or to distinguish between different-sized registers of signs.

G8. There is never complete correspondence between the numeral-signs of a system and the lexical numerals of the language of the society where the system was invented.

G9. There is always some correspondence between the numeral-signs of a system and the lexical numerals of the language of the society where the system was invented.

These two rules obviously complement each other, and at least the first of them has been pointed out by other researchers (Menninger 1969: 53-5). Because I have not discussed lexical numerals in this study, I do not consider them proved, but I know of only one exception to the former and none to the latter. In terms of principle, there are enormous differences between the two representational systems. Most lexical numerals are multiplicative-additive in structure (six **thousand** four **hundred** seven**ty** one), and while some numerical notation systems are multiplicative-additive, such systems are far less common than cumulative-additive, ciphered-additive, or ciphered-positional ones. Numerical notation, is not simply a matter of reducing numerical morphemes into signs. In general, the structural differences between the two systems probably have to do with the fact that lexical numerals are auditory in origin (because they were spoken before they were written), while numerical notation is visual in origin. The one exception is that in some systems of the East Asian family (e.g. the Chinese classical system), numerical notation and lexical numerals are parallel, because the Chinese numerals are signs in the Chinese script, and each character represents one morpheme, so that they can be read directly. As I discussed in Chapter 8, however, this is part of the debate as to whether Chinese numerals are really representations of morphemes or whether they are more like numeral-signs.

In base structure and other signs used to express numeral-phrases, on the other hand, there is frequently a strong correlation between lexical and graphic representations of number. It would be very surprising for a vigesimal numerical notation system to develop among speakers of a language with decimal lexical numerals. Of course, imperialism and other forms of cultural hegemony may spread a numerical notation system far beyond the region of its original invention, and the present worldwide adoption of decimal ciphered-positional numerals indicates that there is little to prevent such numerical notation systems from diffusing to regions whose inhabitants use lexical numerals of various bases. This suggests that, while the initial choice of a base may be determined by the lexical numerals of its inventor(s), once this choice is made, there is relatively little flexibility for changes in base structure.

G10. No system uses numeral-phrases that are read vertically from bottom to top.

This is a near-universal. Numeral-phrases are very often read from left to right, right to left and top to bottom, but the fourth major linear possibility - reading upwards with the highest-valued exponent at the bottom - is extremely rare. The only exception I am able to think of is on Inka *quipus* (ch. 10), where 'top cords' (as opposed to ordinary 'pendant cords') apparently were read in this way (with the highest exponent closest to the main cord). However, since we know so little about how quipus actually were read, this exception may be only an apparent one. This rule is almost certainly related to the fact that it is extremely rare for linear scripts to be read in this direction.

G11. No system uses an identical representation for two different numbers.

This is a near-universal. It is certainly possible to *imagine* a system where the numeral-phrases for 2 and 20 (or any other two numbers) are identical, but this is rarely attested. While such ambiguity is not logically impossible, it creates confusion and reduces the utility of such a system. The converse of this principle is not true: many systems use two or more representations for one number (e.g. Roman VIIII or IX for 9), but this procedure never creates a numeral-phrase whose value is truly indeterminate. This principle does not exclude the possibility that there may be a single *numeral-sign* for two numbers. For instance, the Palmyrene system (ch. 3) uses a single sign for 10 and 100 ( ), but the sign for 100 is always found in conjunction with one or more multiplicative signs, whereas the sign for 10 occurs as part of the cumulative-additive component of the system. Some of the proto-cuneiform signs (ch. 7) have multiple values, but these occur in different sub-systems representing different things; any ambiguity rests solely in modern scholars' failure to recognize the use of different systems in the proto-cuneiform tablets. A true exception to this rule is found in the Sumerian cuneiform system (ch. 7), which uses a vertical wedge ( ) for both 1 and 60. The Sumerians were aware of the ambiguities this could cause, however, and by the Ur III period (2150 to 2000 BC), a different sign for 60 ( ) was used in cases where confusion could result. A similar issue arose in the related Old Babylonian positional cuneiform numerals used in mathematical texts (prior to the invention of zero in the Seleucid period). Except where numbers were lined up in positional columns, any numeral-phrase could have an infinite number of interpretations, though in practice the correct one was often evident from the context.

Implicational Regularities

I1. If a system has a sub-base, the sub-base will always be a divisor of the primary base.

This is a universal. While it is easy to imagine a system with a base of 10 and a sub-base of 3, and such a system would be able to express every number uniquely (and in fact, somewhat more concisely than with a sub-base of 5), this and similar non-divisor sub-bases are never attested in numerical notation systems. Greenberg (1978: 270) notes that at least two lexical numeral systems, Coahuilteco and Sora, have this feature, but it is extremely rare in lexical numerals as well. Of course, numerical notation systems are sometimes structured by means of additional numbers that are not divisors of their primary bases (such as the use of a sign for 4 in the base-10 Nabataean and Kharoshthi systems). These additional numbers are not sub-bases, because they do not recur throughout the system in the way that the sub-base of 5 recurs in Roman numerals (V, L, D).

I2. No ciphered system has a sub-base.

This rule is nearly exceptionless. Of the 23 systems I have studied that have sub-bases, 15 are cumulative-additive, 5 are cumulative-positional, and 3 are multiplicative-additive, but none are ciphered-positional or ciphered-additive. That this should be so is unsurprising; ciphered systems require only one sign per exponent of the base, so introducing a sub-base does not reduce the number of signs required to write a number, as it does in cumulative systems. Yet it is not logically inconceivable for such a system to exist, because a single ciphered sign may be composed of two or more components, and doing so might eliminate the need to develop new signs for higher numbers. There are traces of a sub-base of 10 in the base-20 ciphered Maya head-glyphs (ch. 9), since the signs for 14-19 (and sometimes 13) are expressed by combining the 'bared jawbone' element for 10 with the rest of the sign for the appropriate unit; thus, one need not develop distinct signs for these numbers. However, because it is not used for 11 or 12, and rarely for 13, this is not a full exception to this rule. Apparently, it is not extremely advantageous in most cases to introduce a sub-base solely to obviate the need to develop new signs. We could certainly avoid using the signs for 6 to 9 by introducing a sub-base of 5 into the Western numerals (for instance, if we used a horizontal line $\overline{\ }$ for 5 and $\bar{1}$ $\bar{2}$ $\bar{3}$ $\bar{4}$ in place of 6, 7, 8, and 9), but this might not be particularly useful. We do not seem to have difficulty remembering the ten digits we have.

I3. If a system is cumulative, it will group intraexponential signs in groups of between 3 and 5 signs.

While this regularity, which I call the *Rule of Four*, has a few minor exceptions, it is widespread and very important. Humans are limited in their cognitive capacities, and work most efficiently when information is packaged in groups of three to five bits. Cumulative systems cope with this limit either by using sub-bases (e.g. Roman, Maya bar-and-dot), by using spacing to distinguish groups (e.g. Egyptian hieroglyphs, Aztec), or both (Babylonian sexagesimal). Probably the limit of five is even a bit too high for the human mind to grasp. Two full exceptions to this rule are the Inca *quipu* and the Bambara numerals (ch. 10), which always use groups of up to 9 unit-signs. Partial exceptions include the Hittite hieroglyphs (ch. 2) and the South Arabian numerals (ch. 4), in which chunking in groups of three to five signs was an option, but in other cases groups of up to nine signs were used. I will discuss the origin and effects of the Rule of Four further below ("Subitizing and Chunking").

I4. If a system is multiplicative-additive for a given exponent of its base, it will also be multiplicative-additive for all higher exponents of the base.

I5. If a system is non-multiplicative for a given exponent of its base, it will be non-multiplicative for all lower exponents of the base.

These two complementary rules are exceptionless, and apply to hybrid systems – those that are cumulative-additive or ciphered-additive for some exponents and multiplicative-additive for others. In such systems, it is always the higher rather than the lower exponents that are multiplicative, and once the 'switch' to the multiplicative principle has been made, it applies for all higher exponents. No system is multiplicative-additive for lower exponents and follows some other principle for higher exponents, even though such a system would be workable in theory. The use of hybrid multiplication is primarily useful for extending a system further without the need to develop increasingly large inventories of signs, but comes at the expense of somewhat longer numeral-phrases. It is thus more useful for large exponents than it is for small ones. Moreover, in many lexical numeral systems, multiples of lower exponents are expressed with a single word, but multiples of higher exponents separate the units and exponent components (e.g. Latin *sex, sexaginta, sescenti* vs. *sex milia*). The point at which this shift in lexical numerals occurs sometimes may have affected whether or not a certain exponent is expressed multiplicatively in the corresponding numerical notation system, but it cannot have been the only factor because many systems (e.g. Roman numerals) do not use multiplication at all.

I6. Whenever the multiplicative principle is used in a system, the unit-sign or signs (multiplier) will precede the exponent-sign (multiplicand).

This is a nearly universal rule. It is rarely permitted to express 300 as "100 3" in a multiplicative-additive system; regardless of the base of the system or other structural features, the units precede the exponent. Because of the Ordering Principle, expressions where the exponent-sign was placed first could be interpreted additively in some systems, thereby creating ambiguity, whereas there is no such risk where the unit-signs are placed first. Most lexical numeral systems also place unit-signs first, and thus it is easier to translate a numeral-phrase into its lexical equivalent if this rule is followed in numerical notation as well. An exception to this rule is that in some alphabetic numeral systems, such as the Greek, Coptic, and Cyrillic systems (ch. 5), a small diacritic mark to the left of (before) a sign indicates multiplication by 1000 (e.g. Greek $͵\bar{\gamma}$ = 3000). Such exceptions are possible because in this case there is no possibility of confusing this multiplicative expression with an additive one.

I7. No multiplicative system uses 1 as an exponent-sign.

This rule is virtually exceptionless. Multiplicative-additive systems, by definition, combine unit-sign multipliers with exponent-sign multiplicands, but there is never a separate exponent-sign for the units. Rather, the numbers from 1 up to the base of the system are expressed through unit-signs alone. While an exponent-sign for 1 would be consistent with the principle of combining unit-signs with exponent-signs, it would be completely extraneous and provide no additional information to the reader. While most lexical numeral systems are multiplicative-additive in structure, they also do not use exponent-signs for 1. The only exception to this rule is that the earliest Bamum numerals (ch. 10) apparently had a separate exponent-sign for 1 (which was entirely distinct from the unit-sign for 1). This sign was used only for a brief period (roughly 25 years) before the Bamum system became ciphered-positional. In any event, the use of separate signs for the unit-sign 1 and the exponent-sign 1 prevented any ambiguity from arising. In at least eight other independent cases, there is no exponent-sign for 1: Shirakatsi's Armenian notation (ch. 5), Âryabhata's notation (ch. 6), the Tamil and Malayalam numerals (ch. 6), in Shang China (ch. 8), the Texcocan Kingsborough Codex numerals (ch. 9), Mende (ch. 10), and Pahawh Hmong (ch. 10).

I8. All multiplicative expressions involve only bases or their exponents as multiplicands.

This is a universal. No system uses multiplication involving sub-bases, multiples of exponents of bases, or other additional structuring numbers. It would certainly be possible to have, say, a base-20 system with a sub-base of 5 in which 13 is written as (2x5)+3, but pure addition (5+5+3) is always preferred in such circumstances. Similarly, a decimal system that combines multipliers with 20 (as in the French lexical numeral *quatre-vingt*) is never attested in numerical notation. Complying with this rule helps readers of multiplicative numeral-phrases to distinguish operations involving multiplication from those involving addition.

I9. All composite multiplicands are strictly multiplicative.

This rule is complementary to the previous one, and is also exceptionless. Various multiplicative-additive systems use more than one exponent-sign in combination with a unit-sign multiplier; for instance, the Chinese classical system often uses 萬萬 to write 100 million (10,000 x 10,000), and the Tamil system uses various combinations of ௰ (10), ௱ (100), and ௲ (1000) instead of developing new signs for 10,000 and higher exponents. Doing so is very important, as it is the only way to make a multiplicative-additive system infinitely extendable (see below, rule N1). Rule I8 could theoretically allow *additive* combinations of these exponent-multiplicands - the Tamil numeral ௫௲௲ could be read as 5 x (1000+1000) rather than 5 x (1000 x 1000) - but this never occurs. These 'composite' multiplicands are always multiplicative, never additive or subtractive.

Non-Universals

While the search for cross-cultural universals is important, it can (and often does) go too far, postulating that a given regularity is a universal when in reality it is not. The following regularities are non-universals whose interest lies not in their regularity but rather in the assumption that they are universal, when in fact they have numerous significant exceptions. I held many of these propositions to be universal at the outset of this project, based on my intuition or preliminary reading, but under more careful scrutiny they have proved to be less regular than they first appeared. The existence of these exceptions does not make the generalization irrelevant, but it does require that we take account of the processes that lead to the exceptions and recognize why they are important. I include them as a separate category because their theoretical importance is determined by the fact that they have significant exceptions rather than that they are cross-culturally regular. In all other respects, they are ordinary statistical regularities.

N1. Some additive numerical notation systems are infinitely extendable without the need to invent new signs.

One of the primary benefits cited for positional notation is the fact that any number, no matter how high, can be written with it. However, a few additive systems that use multiplication, such as the Ethiopian numerals (ch. 5), the Armenian numerals of Shirakatsi (ch. 5), Tamil / Malayalam numerals (ch. 6), the Chinese classical numerals (ch. 8), and the Mende numerals (ch. 10), are also infinitely extendable by virtue of their use of repeated exponent-signs as multiplicands. This may also be true of some of the multiplicative techniques employed by Hellenistic mathematicians to overcome the lack of expressions for very high numbers in the hybrid ciphered-additive / multiplicative-additive Greek alphabetic numerals (Heath 1921: 39-41). These techniques obviate the need to develop new signs for higher exponents of the base, and thus produce an infinitely extendable system. In some cases, it is far more concise and expedient to use these systems for writing large numbers than it is to use ciphered-positional systems; for instance, the Ethiopian expression for 100,000,000, ፼፼, requires only two signs where nine Western numerals are needed. It is fair to say that all systems that are *purely* cumulative-additive or purely ciphered-additive are finite in scope.

N2. Some positional systems are not infinitely extendable and hence able to express any natural number.

This is *not* a universal, even though it borders on being a logical necessity. The Cistercian system (ch. 10), which is best understood as one based on orientational position, is clearly not infinitely extendable; once the four positions are occupied (used for writing ones, tens, hundreds, and thousands), the system has reached its end. The same can be said for the Texcocan numerals (ch. 9), which also use orientational position in the form of vertical and horizontal registers. Positional systems that are linear are all infinitely extendable, as one can keep adding new positions in front of the highest exponent. Even in such systems, however, there is always a pragmatic limit based on considerations of conciseness (for instance, one does not often encounter the number 10^{23} written in full positional notation).

N3. Some additive systems use a sign for zero.

It is often thought that systems that have a sign to indicate an empty place must, of necessity, be positional. However, the quasi-positional Maya bar-and-dot numerals (ch. 9) use the zero-sign simply to indicate the absence of a numerical coefficient for time-periods. One of the major reasons why the quasi-positional nature of the Maya numerals has not been recognized is the fact that it very clearly has a sign approximating the role of 0. Furthermore, the multiplicative-additive Chinese classical and commercial systems (ch. 8) use signs for zero to indicate blank positions even though they are non-positional and thus do not need to do so, strictly speaking. The zero in all of these cases adds redundancy to the system, which may serve to clarify the meaning of a phrase. This function of zero is quite distinct from that in positional systems, where it is used to specify the position of non-zero digits, and thus to identify the exponent by which they should be multiplied.

N4. Some systems are not written and read in a one-directional straight line.

The vast majority of numerical notation systems are purely linear, whether they are read from left to right, right to left, or top to bottom; however, several systems are read in a more convoluted direction. The Chinese commercial system (ch. 8) and the Texcocan numerals of the Kingsborough Codex (ch. 9), both of which are multiplicative-additive, place the unit-signs in a row beneath the corresponding exponent-signs, so that the phrase is written and read in a zigzag fashion. The same is true for the multiplicative component of the Greek alphabetic numerals (ch. 5) above 10,000, except that the unit-signs are in this case placed above the multiplicative sign for 10,000 (M). Other alphabetic systems that use multiplication often do so by placing a stroke above or below another numeral, thus requiring a less-than-linear reading of the numeral-phrase. The most extreme non-linear systems are the Cistercian numerals (ch. 10) and the ordinary Texcocan numerals (ch. 9), both of which use orientational position rather than linear position, the former through four rotational orientations, the latter through horizontal and vertical registers.

N5. Not all independently invented systems are cumulative-additive.

Based on limited evidence from the circum-Mediterranean region, or even based on non-empirical theoretical reasoning, it has often been claimed that cumulative-additive systems are the most ancient or basic form of numerical notation (Hallpike 1979; Damerow 1996; Dehaene 1997). However, the falsity of this assumption is evident from even a cursory examination of the evidence presented in this study. The Shang numerals (ch. 8) are multiplicative-additive, the Inca *quipu* numerals (ch. 10) are cumulative-positional, and the Bambara numerals (ch. 10) are ciphered-additive (with a cumulative component). If it should turn out that the Brāhmī numerals (ch. 6) were developed independently of Egyptian or Greek influence, then we would have an additional example of an independently invented ciphered-additive system. While it is possible that these non-cumulative-additive systems had cumulative-additive antecedents for which no evidence survives, this seems unlikely given our present state of knowledge. It is probably not coincidental that several independently invented systems of antiquity (Egyptian hieroglyphs, proto-cuneiform, Indus) and others that may be independently invented (Etruscan) are cumulative-additive, because the widespread existence of cumulative unstructured tally-marks in non-literate civilizations is an obvious antecedent to numerical notation. Yet this tendency does not reach the status of a universal evolutionary law or even a solid generalization.

Cognitive Explanations of Synchronic Regularities

In Chapter 1, I briefly mentioned various aspects of cognition relating to numeration, and suggested that it might be productive to examine the features of numerical notation systems that are relevant to their usefulness for various functions. The primary function of numerical notation is always the simple visual representation and reading of numbers, and only rarely or secondarily arithmetical computation. However, consideration of these factors in the abstract is not likely to be very useful, and thus I have left their analysis until now, so that I can provide relevant examples. To explain synchronic regularities in the structure of systems involves comparing them with certain cognitive criteria, thereby establishing potential reasons why they should exist. Synchronic regularities are features that are true of all or most systems, or, in the case of implicational regularities, are true of all or most systems that possess a certain feature. This presents a slight difficulty, because one cannot explain an exceptionless rule by examining the circumstances in which it does or does not apply. However, because these rules are not logical requirements, it is possible to imagine systems that violate them; hence this difficulty can be overcome by considering what would result if they were violated and seeing how certain cognitive conveniences are satisfied by conforming to them. Where there are actual exceptions to these regularities, this task can be even more enlightening than when exceptions are merely hypothetical.

I recognize that, in a sense, this is an indirect form of explanation; ideally, we would like to have more information about the decision-making processes and behaviour that produced these regularities. However, since individual inventors were very rarely considerate enough to have left detailed records concerning their choice of a specific principle or base, and because very few systems were invented within living memory, indirect techniques are necessary. Moreover, since there is no guarantee that individuals were consciously aware of the specific advantages of different means of representation, often it may be more productive to analyse representational efficiency without reference to individual decision-making. I believe that the following factors are always relevant *to some unknown degree* in the decision-making processes relating to the development and use of numerical notation systems. It is simply impossible that so many specific regularities in the structure of numerical notation systems - ones that are not strictly determined by the logic of those systems - would emerge unless the various representational techniques were constrained by certain cognitive considerations. This does not mean, however, that cognitive factors provide *full* explanations for the observed phenomena. It remains to be evaluated exactly how important they are with respect to one another and in comparison to social factors.

Phrase Ordering

One of the central principles of the cognitive sciences (of which cognitive anthropology is an important branch) is that information is often most useful when it is structured. Since numerical notation is a representational system used to help record, remember, and use numerical information, it should come as no surprise that several of the synchronic regularities I detailed above relate directly to the ordering of numeral-signs within numeral-phrases. Most obvious among these is the Ordering Principle (G5), which orients all systems in a highest-to-lowest direction of exponents. Yet a number of other rules, including G6 (absence of signs for operations), G4 (absence of division as an operation), I4 and I5 (governing the switch from addition to multiplication in hybrid systems), I6 (unit-signs precede exponent-signs), and I7 (1 is not an exponent-sign), relate more or less directly to the arrangement of numeral-signs within numeral-phrases. That so many regularities pertain to the ordering of numeral-phrases suggests that something important is going on, from a cognitive perspective, to constrain the order of signs that is possible in numerical notation systems.

The logical place to start in the analysis of the Ordering Principle is to look at similar attempts to explain ordering in lexical numerals by linguists working within a generative-transformational framework (Salzmann 1950; Hurford 1975, 1987; Stampe 1976). Yet, while this research is important as a description of systems, it cannot do much to explain the phenomena it describes, unless the assumption is made that descriptive rules map perfectly onto cognitive processes. While I have called the regularities that I have described above 'rules', I do not mean to assert that these are applied by individuals, either consciously or unconsciously. In most cases, I think they are the outcomes of broader cognitive principles that relate to the structuring of numerical information.

In positional systems, signs must be put in their proper order to ensure that a numeral-phrase is interpreted correctly. A positional system that did not do this would be completely unworkable, since every numeral-phrase would have many equally valid readings. In many other systems, however, there is no logical requirement prohibiting irregular ordering. In ordinary cumulative-additive and ciphered-additive systems, for example, the signs can be placed in any order without any ambiguity, since the values of the signs are simply added. Moreover, in multiplicative-additive systems, as long as each unit-sign is associated with a specific exponent-sign, the resulting sign pairs can be placed in any order. Yet the Ordering Principle is nearly exceptionless, so such irregular phrases are rarely acceptable, and occur only where the writer has made an error.

In cumulative systems, a certain degree of ordering is necessary to ensure that identical signs are grouped together. If 327 could be written in Roman numerals as ICIXCVCX, the advantage of cumulation (the adding of identical signs) would be greatly reduced by the fact that identical signs are far apart from one another. Even with the requirement that the signs for each exponent be grouped together, one could still write CCCVIIXX, XXCCCVII, XXVIICCC, VIICCCXX, or VIIXXCCC instead of CCCXXVII, the only acceptable form. The Ordering Principle applies to all systems, and regulates intraexponential ordering in cumulative systems as well as interexponential ordering in both additive and positional systems.

Greenberg (1978: 274) suggests that one cognitive principle favouring "larger + smaller" formations in lexical numerals is that, by beginning with the largest exponent, the first element closely approximates the final result, producing an expectation of the eventual size of the phrase in the listener / hearer. I think that explanations involving this factor of 'successive approximation' also apply to numerical notation, and that the desire rapidly to approximate a value is in part responsible for the Ordering Principle. In particular, it explains why numeral-phrases that are in order, but from the *lowest to highest* exponent, never occur. The only difference that might be relevant is that, because it is a visual medium, numeral-phrases could theoretically be read in any order, regardless of the manner in which they are written, whereas spoken lexical numerals obviously must be heard sequentially.

Using successive approximation as the only explanatory factor leaves the tricky problem that, while numerical notation systems are nearly always ordered from highest to lowest exponents, lexical numeral systems are not. The few exceptions to the Ordering Principle in numerical notation (e.g. Greek, Cyrillic, and Glagolitic alphabetic numerals for 11-19) are a direct result of a comparable irregular ordering in the corresponding lexical numerals. We then might expect to find violations of the Ordering Principle wherever the lexical numerals of a system's users also do so. Since 17 in Latin is *septendecim,* we should expect the Roman numeral for 17 to be VIIX, which of course is unacceptable. Upon reading the number 16 aloud as *sixteen,* English speakers rapidly transform the 'high-low' order numeral-phrase into a 'low-high' lexical numeral. The explanation for why numerical notation is ordered so strictly must be somewhat different from explanations regarding the ordering of lexical numerals.

In many circumstances, ordering constraints in numerical notation are a consequence of the omnipresent concern with avoiding ambiguity, coupled with the relative inflexibility of signs within numerical notation systems. One of the primary features of almost all numerical notation systems is that they are designed to minimize the possibility that a reader will misinterpret a given sign or series of signs; that is, each series of signs has only one numerical meaning (rule G11). Lexical numeral systems can use a variety of techniques other than ordering to eliminate ambiguity. For instance, modern German (among many other European languages) uses stem alteration to distinguish 16 (*sechzehn* = 6+10) from 60 (*sechzig* = 6 x 10), and Classical Sanskrit uses pitch accent alone to distinguish 108 (*aštáçatam* = 8+100) from 800 (*aštaçatám* = 8 x 100), even though the numerical value and ordering of the two elements in each word are identical. Numerical notation systems do not have this flexibility; one of their conveniences is that they use a relatively limited set of discrete and inflexible signs. Therefore, the strict ordering of numeral-phrases is one of the only ways to assign values unambiguously to signs.

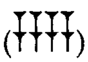
Ordering is also essential for unambiguously indicating which arithmetical operations are to be used in combining numeral-signs to derive the values of numeral-phrases. Numerical notation systems do not explicitly express signs for the operations being used (rule G6), and thus in any system, using the order of signs is essential to identify the operations used to obtain the final numerical value. The rule that unit-signs precede rather than follow exponent-signs in multiplicative systems (I6) is designed to specify that the two values are to be multiplied rather than added, allowing the unambiguous reading of numeral-phrases without the need for signs for operations. If it did not apply, and 300 could be written as '100 3', then the numeral-phrase could be interpreted multiplicatively (as 100x3, or 300) or additively (as 100+3, or 103).

No one of these factors is sufficient to explain why the Ordering Principle is so prevalent in numerical notation systems worldwide. I suspect that the avoidance of ambiguity is foremost among these (being essential to ordering in positional systems and important to many additive ones), with the principle of successive approximation being also very important. In most cases, this will lead to considerable conformity in phrase-ordering with lexical numerals, but this cannot be of primary importance because complete correspondence in ordering between numerical notation and lexical numerals appears to be rare (cf. rule G8).

Subitizing and Chunking

The Rule of Four (rule I4, above) is a nearly exceptionless regularity in cumulative systems that governs the widespread use of grouping signs in three to five (4 ± 1) units, tending significantly towards the lower end of this range. This feature developed independently at least seven times (Egyptian hieroglyphic, Etruscan, proto-cuneiform, Chinese rod-numerals, Maya bar and dot, Indus Valley, Iñupiaq) and, with very few exceptions, has been used in all cumulative systems throughout history.

In general, the direct causative role of basic cognitive principles on the structure and evolution of numerical notation systems is limited. Yet the existence of the Rule of Four is explained most parsimoniously by reference to the process of subitizing, or the ability to enumerate rapidly small quantities of discrete objects without having to count them explicitly. As discussed in Chapter 1, humans have been shown experimentally to be able to enumerate groups of between one and three dots rapidly and with almost no error, and groups of four dots with some error and slightly less quickly, but most individuals cannot count groups of five or more dots without significant error or considerable delay (Mandler and Shebo 1982). While the origins of subitizing are still unclear, a reasonable working hypothesis is that it results from the physiological constraints of the mechanism by which our visual system localizes objects in space (Dehaene 1997: 68). The implications of this principle for numerical notation are obvious: long groups of undivided cumulative signs (e.g. IIIIIII, IIIIIIII, IIIIIIIII) will take longer to read and result in more errors than if some technique is used to avoid them.

A common way in which systems conform to the Rule of Four is by dividing long sets of signs into smaller groups through spacing. Cumulative systems that do not use 5 as a sub-base, including most of the systems of the Hieroglyphic, Levantine, and Mesopotamian families, as well as the Indus numerals, divide long sets of identical signs either by placing groups of three to five signs side by side () or above one another (). This operation has no effect on a system's conciseness, sign-count, or extendability, but it provides an advantage to numerical notation systems that do so over those that do not - they are more easily readable by individuals, who are constrained by human limits on the capacity to process bits of information.

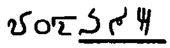
The use of a sub-base, found in the Italic family, in many systems of the Mesopotamian and Mesoamerican families, the Chinese rod-numerals, and the Iñupiaq numerals, is another technique to allow a system to conform to the Rule of Four. In these systems, instead of using multiple groupings of three to five signs within each exponent, the use of a sub-base means that there is never any need to use more than four signs of any one type (e.g. VIIII instead of III III III). Where the system's primary base is 20 (Mesoamerican and Iñupiaq), not only does a sub-base of 5 ensure that the sign for 1 need only be repeated up to four times, but the sign for 5 need only be repeated up to three times, again in conformity with the Rule of Four. Finally, in the base-60 systems of Mesopotamia, the use of a sub-base of 10 ensures that the sign for 10 never needs to be repeated more than five times, since 60 is not represented with six signs for 10 but with a single sign for 60.

Even in systems that lack a true sub-base, such as the members of the Levantine family, the Rule of Four is brought to bear on the system's structure. These systems all have a base of 10, but many also use additional structuring signs – always including 20, and sometimes also 4 and 5 (excepting Phoenician and Aramaic). By using signs for 4 (Kharosthi and Nabataean) and 5 (Hatran, Palmyrene, Syriac, and Nabataean), the sign for 1 need never be repeated more than three or four times. The additional sign for 20 also helps establish the Rule of Four, because it need only be repeated four times at most in writing any number up to 100. Finally, because the Levantine systems are multiplicative above 100, the same principles used to write numbers less than 100 allow any number up to 1000 (most of them go no higher) to be written without the need to repeat any sign more than four times.

Further confirmation that the Rule of Four is a consequence of subitizing is found in diachronic changes in systems in which techniques are introduced to reduce phrases that had four or five repeated signs to ones that only needed three or four repeated signs. The republican Roman numerals (ch. 4) were purely additive, and required up to four cumulative signs for each exponent of the base, but in the late republican period, the introduction of subtractive notation for 4 and 9 meant that a writer had the option of using phrasing that required only three signs of each type at most (Sandys 1919). In the Sumerian cuneiform numerals (ch. 7), the numbers 7, 8, and 9 were written as 𒐌, 𒐍, and 𒐎, respectively. When this system was adopted and modified into the Assyro-Babylonian common system and the Babylonian positional system, these cumulative phrases were altered into forms that grouped signs in sets of no more than three signs (𒐌, 𒐍, and 𒐎). Finally, the early Chinese rod-numerals (ch. 8) expressed 4, 5, and 9 as 𝍣, 𝍤, and 𝍨, which require 4 or 5 repetitions of single signs.[4] This was necessary, because the system's structure was partly a consequence of the use of physical rods as computational tools. However, in the Song dynasty, when written rod-numerals were used extensively in mathematical texts, the older, purely cumulative forms were sometimes replaced with ciphered signs: 〤, 〥, and 〩, so that no phrase required more than three repeated signs. These three independent reductions in the number of repeated signs required strongly suggest that five signs is cognitively too many, and that even four signs may be difficult to perceive.

[4] I have listed the *zong* forms only, but the *heng* forms are simply 90° rotations of the former and thus the same principle applies.

In non-cumulative systems, signs are not repeated and do not need to be counted, and thus subitizing is irrelevant. A different set of cognitive principles related to the processing of visual information applies to ciphered and multiplicative systems, namely chunking. As first described by Miller (1956), humans have a limited ability to memorize and recall long sets of bits of information, with the maximum being the "magical number" 7±2. In order to deal with long lists of information, it is much easier to *recode* input into a series of chunks, each of which contains a small number of bits. In practice, chunks of three or four bits produce an effective balance between the limits of the memory, which restricts the maximum size of chunks, and the desire to minimize the number of chunks necessary to recode input.

Chunking has a considerable effect on the structure of non-cumulative numerical notation systems. For instance, in Western numerals and many other ciphered-positional systems, it is typical to divide long numbers up into sets of three numbers (e.g. 123,456,789). Four-digit numbers are sometimes but not always grouped in this way (1000 vs. 1,000), but it is normal for all five-digit and longer numbers to be subdivided. Doing so not only groups large series of numbers into manageable chunks, it also accords precisely with the millesimal (base-1000) structure of American English lexical numerals (thousand, million, billion, trillion, etc.), and to a lesser degree with the mixed base-1000/base-1,000,000 lexical numerals of British English and many other European languages, which are themselves probably governed by chunking. The same reasoning applies to the special role accorded to 10,000 (and later to 100 million) in the otherwise decimal Chinese classical numerals. In this case, the system is multiplicative-additive and does not use additional signs to divide numeral-phrases into chunks, but the same principle applies, so that there are four pairs of unit-signs plus exponent-signs in each chunk. Finally, I propose that one of the reasons that most hybrid ciphered-additive / multiplicative-additive systems switch to multiplication at 1000 or 10,000 is that doing so groups signs into chunks of no more than three and four signs, respectively. For instance, in the Fez numerals (ch. 5), which use a multiplier at 1000 by placing a horizontal line under a sign or signs, 658,379 is written (reading from right to left) as  , or (9+70+300) + (8+50+600) x1000, thus dividing the numeral-phrase into two chunks of three bits on the basis of the subscript multiplier used. I think it likely, however, that chunking in numerical notation in all of these cases is probably a consequence of the prior chunking of their lexical numerals.

Whether there is a connection between subitizing and chunking is a question for cognitive psychologists, and cannot be resolved here. Subitizing may in fact be a specific example of how chunking affects humans' ability to perceive and encode information, one that is restricted to enumerating patterns of discrete visual objects. Chunking has a much broader range of applications, as it is not restricted to visual information and applies to tasks other than simple enumeration. Subitizing, as the direct cause of the Rule of Four, has far more significant effects on numerical notation systems than does chunking in general.

Fingers and Numbers

The Rule of Ten (rule G2, above) is an exceptionless rule that all systems have 10 or a multiple of 10 as their primary base. Approximately 90% of all systems have 10 as their primary base, with 20 being the next most frequent at about 7%, while three systems (proto-cuneiform, Sumerian, Babylonian positional) have a primary base of 60 and one (Âryabhata's numerals) a primary base of 100. Why should this be? In the pseudo-Aristotelian *Problemata* (Book XV.3, 910 b23 - 911 a4), the question is posed, "Why do all men, whether barbarians or Greeks, count up to ten, and not up to any other number ... It cannot have been chance; for chance will not account for the same thing being done always: what is always and universally done is not due to chance but to some natural cause" (Heath 1921: 26-27). After discarding several fanciful suggestions, the author finally asks, "Or is it because men were born with ten fingers and so, because they possess the equivalent of pebbles to the number of their own fingers, come to use this number for counting everything else as well?"

While this particular passage refers specifically to lexical numerals rather than numerical notation, it will no doubt occur to any reader that numerical notation systems may tend to be decimal because humans have ten fingers. However, this is too simplistic. I contend that, while the ultimate cause of decimal numeration may be that we have ten fingers, the proximate cause is that the vast majority of the world's languages have decimal lexical numerals. Where this is not the case, as in Mesoamerica (base-20) and early Mesopotamia (base-60), the numerical notation systems that develop are non-decimal, though of course they still comply with the Rule of Ten. Wherever numerical notation develops independently, the system that is developed has the same primary base as its inventors' lexical numerals (see rule G9). The existence of non-decimal numerical notation systems refutes any simple causal relation between fingers and numerical notation; thus, the question posed in the *Problemata* needs to be restated to take account of the fact that decimal lexical numeration is not universal. The evidence suggests an overwhelming influence of lexical numerals on the initial choice of base of a numerical notation system, which may occur millennia after the development of a numerical base in a language's lexical numerals. While lexical numerals are constrained by the nature of the hands, they are not determined by them, as seen in the host of non-decimal (and even non-base-structured) lexical numerals in the world's languages.

Another feature of numerical notation systems may be explained in part by considerations related to the fingers, namely the development of a sub-base of 5 in at least three independent cases: the Etruscan numerals (ch. 4), Chinese rod-numerals (ch. 8), and the Mesoamerican bar and dot numerals (ch. 9).[5] The existence of five handy

[5] A fourth possible instance is the development of a special sign for 5 in some early Aramaic inscriptions (ch. 3); however, it seems equally possible that this development (which was never

cumulative-like digits on the end of each hand is too obvious a coincidence to overlook. Moreover, in one unusual case, the Iñupiaq numerals, we know that one of the stimuli to which its student inventors had been exposed was Chisanbop finger-computation. Again, however, it is worthwhile to look to the lexical numerals of these regions for other possible explanations. The Iñupiaq lexical numerals have a quinary sub-base, which was part of the reasoning used by the students in designing their system. Moreover, it is probable that there was a quinary component to the lexical numerals of the inventors of the Mesoamerican bar and dot numerals, who were probably Zapotec or Mixe-Zoquean speakers (Colville 1985: 796). In these cases, it is more parsimonious to presume that the existence of a lexical sub-base of 5 partly inspired the similar graphic sub-base. Nevertheless, the Etruscan lexical numerals probably had no quinary component and the Chinese numerals certainly did not. This raises the possibility that the fact that there are five fingers on each hand is directly related to the use of quinary sub-bases in such systems. There is another possibility to be considered, which is that in a system with a sub-base of five, no one sign will need to be repeated more than four times, thus enabling a system to conform with the Rule of Four (rule I4), as discussed above. The fact that both the Etruscan numerals and Chinese rod-numerals were used very early in the context of arithmetical computation employing physical counters is further evidence of the need to keep the number of repeated signs to a minimum. 5 is the only reasonable choice for a sub-base for a decimal system because of rule I2, which states that sub-bases must divide evenly into bases. Yet this merely extends further the causal chain in these cases: ten fingers lead to decimal lexical numerals, which lead to decimal numerical notation, which then lead – in combination with the Rule of Four - to quinary sub-bases in numerical notation.

---

extended into a full sub-base) was a result of contact with the Italic family of systems (ch. 4). The Ryukyu *sho-chu-ma* numerals (ch. 8) are probably derived in part from the rod-numerals, and the Zuñi numerals (ch. 10) were probably borrowed from the Roman numerals.

In summary, whether we are considering the origins of decimal primary bases or quinary sub-bases in numerical notation systems, the direct role of the fingers is not as great as might be thought, although their indirect influence cannot be denied. The particular effects of various factors, including - but possibly not limited to - lexical numerals, the fingers, and chunking of visual information, are apparently complex, and we will likely never understand the causal relations precisely. A fuller examination of the bases of the lexical numeral systems of ancient civilizations is an important topic for future study.

Diachronic Regularities

The synchronic regularities I have just outlined are constraints that govern the limits of variability we expect to find among the world's numerical notation systems (allowing, of course, for a few exceptions here and there). They dictate that some forms are simply not feasible (e.g. a base-16 system with a sub-base of 3 that is ciphered-positional for the first two exponents and multiplicative-additive thereafter). Such systems, though they would fulfil the basic definitional criteria of numerical notation systems, are either very taxing on human cognitive capacities or extremely counter-intuitive. Even with these constraints, there is still considerable flexibility for a variety of systems to exist (otherwise, there would be only one combination of basic principles, not five). Yet this variability is non-random. Strong and important similarities are shared by the systems of the various regional phylogenies that I have dealt with in each separate chapter, although these similarities do not approach the character of cross-cultural regularities. Even more importantly, there are rules that govern how systems change over time.

I will now examine diachronic regularities, which apply not to individual systems, but to temporal trends among systems. These exemplify change rather than stasis in numerical notation systems. To analyze diachronic regularities requires that we shift emphasis by moving away from the numerical notation system as the unit of analysis and towards the *event of transformation* (cf. Mace and Pagel 1994). Two processes of change exhibit diachronic regularities. First, there is the process of **transformation**, in which an older system gives rise to a new one that is structurally different from it. This process presumes a direct phylogenetic relationship between the ancestral and descendant systems, but does not tell us what happens to the ancestor after it gives rise to the descendant. To analyse the transformation of systems, we must establish the structure of both the ancestral system and its descendant, correctly identify that the latter is derived from the former, and ideally determine the nature of the process by which the new system arose from the old. For the purposes of this analysis, 'transformation' does not include cases where the ancestral and descendant systems have the same basic structure. The second process is that of **replacement**, in which one system becomes extinct and is supplanted by another. It does not matter whether the system being replaced is directly related, indirectly related, or unrelated to its successor. Even though there may be no resemblance between ancestor and descendant, it is usually easiest to identify a system that replaces one that goes extinct provided that good regional chronological sequences exist.

I will show that both transformation and replacement are severely constrained in their possible outcomes, and thus, while there are far fewer diachronic regularities than synchronic ones, their effects on the pattern of historical change in numerical notation over time are remarkable. Where diachronic regularities exist, there are non-random patterns of cultural change that meaningfully can be called *evolutionary*. To admit that the cultural evolution of numerical notation is real is not to concede that it is linear, nor does it require that such changes be regarded as adaptive.

Transformation of Systems

Table 11.1 summarizes all the instances covered in this study where a system uses a different intraexponential or interexponential principle than its ancestor. These comprise all cases of transformation of principle for which adequate evidence exists, considering only the five basic principles, but omitting other features (base, use of multiplication for higher exponents, and other structuring signs). I have also omitted cases, such as the Ryukyu *sho-chu-ma* tallies (ch. 8), which are almost certainly different in principle from their ancestor, but whose ancestor cannot be identified clearly. In all, 22 systems use a different principle than their ancestor.

Table 11.1: Transformation of Systems (by case)

Ch	Ancestor(s)	Principle	Descendant	Principle
9	Aztec	Cumulative-additive	Kingsborough codex	Multiplicative-additive
2	Egyptian hieroglyphic	Cumulative-additive	Egyptian hieratic	Ciphered-additive
9	Maya additive	Cumulative-additive	Maya positional	Cumulative-positional
4	Roman	Cumulative-additive	Roman positional	Cumulative-positional
7	Sumerian	Cumulative-additive	Babylonian positional	Cumulative-positional
8	Chinese classical	Multiplicative-additive	Jurchin	Ciphered-additive
8	Chinese classical	Multiplicative-additive	Chinese positional	Ciphered-positional
10	Bamum	Multiplicative-additive	Bamum (mfemfe)	Ciphered-positional
6	Âryabhata	Multiplicative-additive	Katapayadi	Ciphered-positional
6	Malayalam	Multiplicative-additive	Malayalam (modern)	Ciphered-positional
10	Pahawh Hmong	Multiplicative-additive	Hmong (2nd stage)	Ciphered-positional
9	Maya bar and dot	Cumulative-positional	Maya head glyph	Ciphered-positional
5	Greek alphabetic	Ciphered-additive	Greek positional	Ciphered-positional
6	Brāhmī	Ciphered-additive	Indian positional	Ciphered-positional
6	Brāhmī	Ciphered-additive	Tamil / Malayalam	Multiplicative-additive
6	Brāhmī	Ciphered-additive	Âryabhata	Multiplicative-additive
5	Armenian alphabetic	Ciphered-additive	Armenian (Shirakatsi)	Multiplicative-additive
10	Indian positional	Ciphered-positional	Varang Kshiti	Ciphered-additive
10	Western	Ciphered-positional	Cherokee	Ciphered-additive
10	Western	Ciphered-positional	Iñupiaq	Cumulative-positional
10	Western / Arabic	Ciphered-positional	Bamum	Multiplicative-additive
10	Western / Arabic	Ciphered-positional	Mende	Multiplicative-additive

From this, we can see that there is considerable variability in the possible transformations of systems. Table 11.2 quantifies the frequencies of these structural transformations, first by graphing the changes according to both intraexponential and interexponential dimensions, and then by considering each dimension of change separately.

Table 11.2: Transformation of systems (frequency)

Ancestor's Structure	Descendant's Structure: Cu-Ad	Cu-Po	Ci-Ad	Ci-Po	Mu-Ad	Total
Cu-Ad	x	3	1	0	1	**5**
Cu-Po	0	x	0	1	0	**1**
Ci-Ad	0	0	x	2	3	**5**
Ci-Po	0	1	2	x	2	**5**
Mu-Ad	0	0	1	5	x	**6**
Total	**0**	**4**	**4**	**8**	**6**	**22**

Intraexponential Changes		Interexponential Changes	
Cu-Ci	1	**Ad-Po**	10
Ci-Cu	2	**Po-Ad**	4
Cu-Mu	1		
Mu-Cu	0		
Ci-Mu	5		
Mu-Ci	6		

Alternately, these changes can be represented graphically as in Figure 11.1. Vertical arrows indicate intraexponential transformations, horizontal arrows indicate interexponential transformations, and diagonal lines involve both types of change, with the numbers indicating the frequency of each change. Grey lines indicate changes that are only attested in modern contexts (1800 – present).

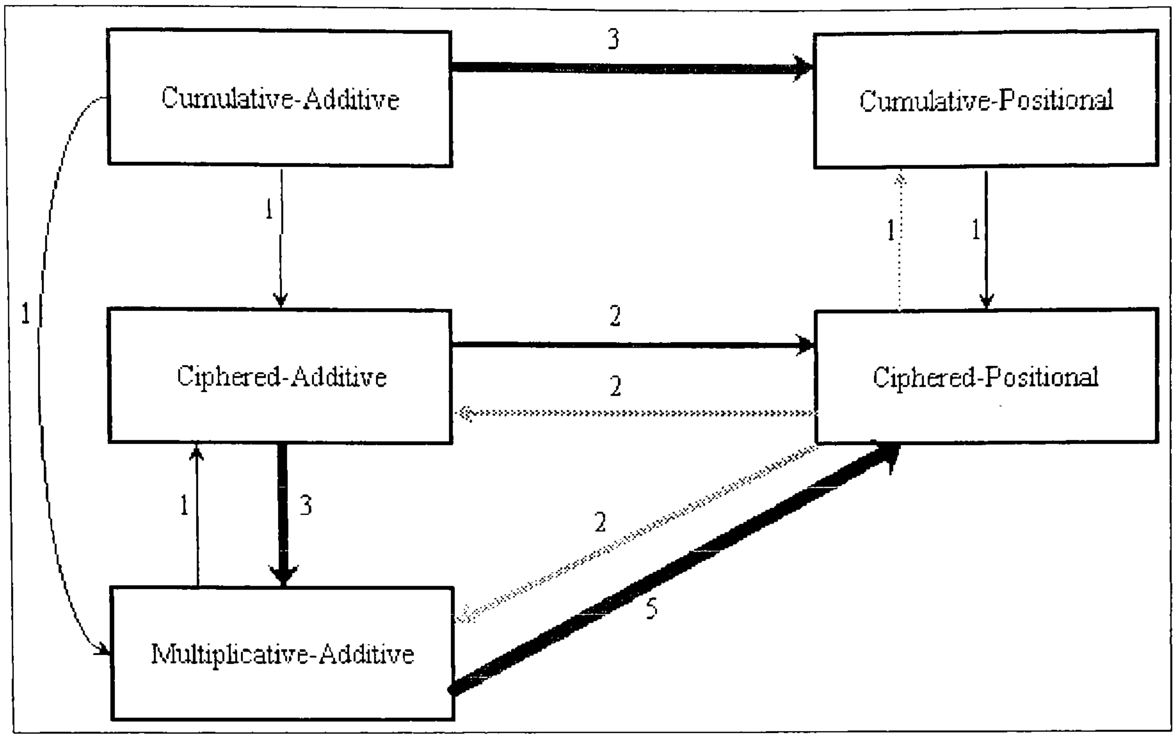


Figure 11.1: Transformations of systems

On the surface, there is enormous variability in the possible transformations, and it thus seems unlikely that any regularities can be extracted by examining the process of the invention of systems. Yet on closer inspection, three important regularities can be extrapolated from these data:

T1. No additive system develops from a positional ancestor. While this is not a universal, it greatly constrains the evolutionary history of numeration. In ten cases throughout this study, additive systems gave rise to positional ones, while in only four cases did the reverse occur. This finding takes on greater importance when we examine the four exceptions to this rule: the ciphered-additive Cherokee and Varang Kshiti systems and the multiplicative-additive Bamum and Mende systems (all Chapter 10).[6] These systems were all developed in the colonial period by inventors whose knowledge of the ciphered-positional antecedents of the systems from which theirs were derived (the Western, Arabic, and Indian systems) was limited. None of these systems has been

[6] If the Ryukyu cumulative-additive numerals were developed on the basis of the Chinese rod-numerals (ch. 8), this would constitute a fifth exception.

notably successful: one (Cherokee) was rejected at the time of its invention and another (Bamum) was transformed by its inventor into a ciphered-positional system within twenty years of its invention. When dealing with the pre-modern development of numerical notation systems, this rule is truly universal; no additive system prior to the nineteenth century had a positional ancestor.

T2. No cumulative system develops from a non-cumulative ancestor. Again, this rule has one exception, that being the development of the Iñupiaq cumulative-positional numerals (ch. 10) on the basis of the Western system. Because this system was developed very recently and in an educational context, it is not clear to what extent this represents a true exception; the numerals' long-term survivability is probably limited. No cumulative-additive system has emerged from any system other than another cumulative-additive one (again, the Ryukyu numerals *may* be an exception). It should be noted that it is rare for any sort of intraexponential transformation to involve cumulative systems (either as ancestor or descendant). The vaunted transformation that occurred when the cumulative-additive Egyptian hieroglyphic numerals were cursively reduced into the ciphered-additive hieratic numerals (ch. 2) is the only Old World example of such a change, with the other two cases being the transformation of the Aztec numerals into the variant multiplicative-additive form seen in the Kingsborough Codex and the invention of the Maya head-glyph forms as alternatives to the older bar and dot numerals (ch. 9).

T3. The only transformation that involves both intra- and interexponential change is the invention of multiplicative-additive systems from ciphered-positional ones, and vice versa. This is an exceptionless rule. Of the nine unattested transformations in Table 11.2 (cells with a 0 value), six involve both an intraexponential and an interexponential change.[7] These changes are presumably too radical alterations of

[7] The other three are all changes in principle resulting in cumulative-additive systems, which as I have already stated, is never known to have occurred.

principle to be likely to occur. Yet the rise of ciphered-positional systems based on multiplicative-additive antecedents occurs five times (albeit sometimes in conjunction with some externally introduced knowledge of positionality). While this transformation involves both intra- and interexponential change, it is nonetheless relatively simple, involving only the elimination of exponent-signs and the addition of a sign for zero. In two other cases (the Bamum and Mende numerals, already mentioned), the reverse change occurs, with ciphered-positional systems giving rise to multiplicative-additive descendants.

In summary, these regularities tend over time to increase the frequency of non-cumulative systems over cumulative ones, and of positional over additive systems. The reverse changes only occur in modern contexts, and the resulting systems have not been extremely successful. Among non-cumulative systems, there is no trend favouring ciphered over multiplicative systems, or vice versa. Of the twenty possible transformations (omitting 'transformations' where the descendant's structure is identical to the ancestor's), only eleven are attested, and only eight are attested in pre-modern contexts. Three of these transformations (encompassing eight of the 22 examples of change in Table 11.2) result in the creation of ciphered-positional systems. The trend towards ciphered-positional notation is partly explained by this transformational pattern.

A second type of inventive change does not involve changes in intraexponential or interexponential principle, but only to the use of multiplication in higher exponents of a system's base. While only a limited number of systems use such a feature, it produces a regularity that is quite important for understanding the diachronic patterning of systems.

T4. When one system that uses the multiplicative principle gives rise to another, the exponent above which the descendant is multiplicative is never higher than that of the antecedent. Figure 11.2 indicates all the systems that use the multiplicative principle for some exponents and which have a multiplicative ancestor. Many other hybrid multiplicative systems (e.g. Cherokee, South Arabian, Roman multiplicative) have non-multiplicative ancestors, but these are not relevant to this rule. The number in each box indicates the exponent(s) at and above which the multiplicative principle is used (with the number 1 indicating fully multiplicative-additive systems). Ancestral systems normally begin to use multiplication at a point equal to or higher than their descendants. Solid lines indicate cases that obey this rule, while dotted lines indicate exceptions. Both the Syriac and Epakt exceptions are only partial exceptions, since the *lowest* point at which they use multiplication (1000) is equal to that of their ancestors, but they also use a different multiplicative technique for 10,000 and above. Notably, the two remaining exceptions, the Jurchin (ch. 8) and Mari (ch. 7) systems, are the only two cases outside the super-family encompassing the Hieroglyphic, Alphabetic, and South Asian families. This suggests the possibility that this rule may only apply within this larger group.

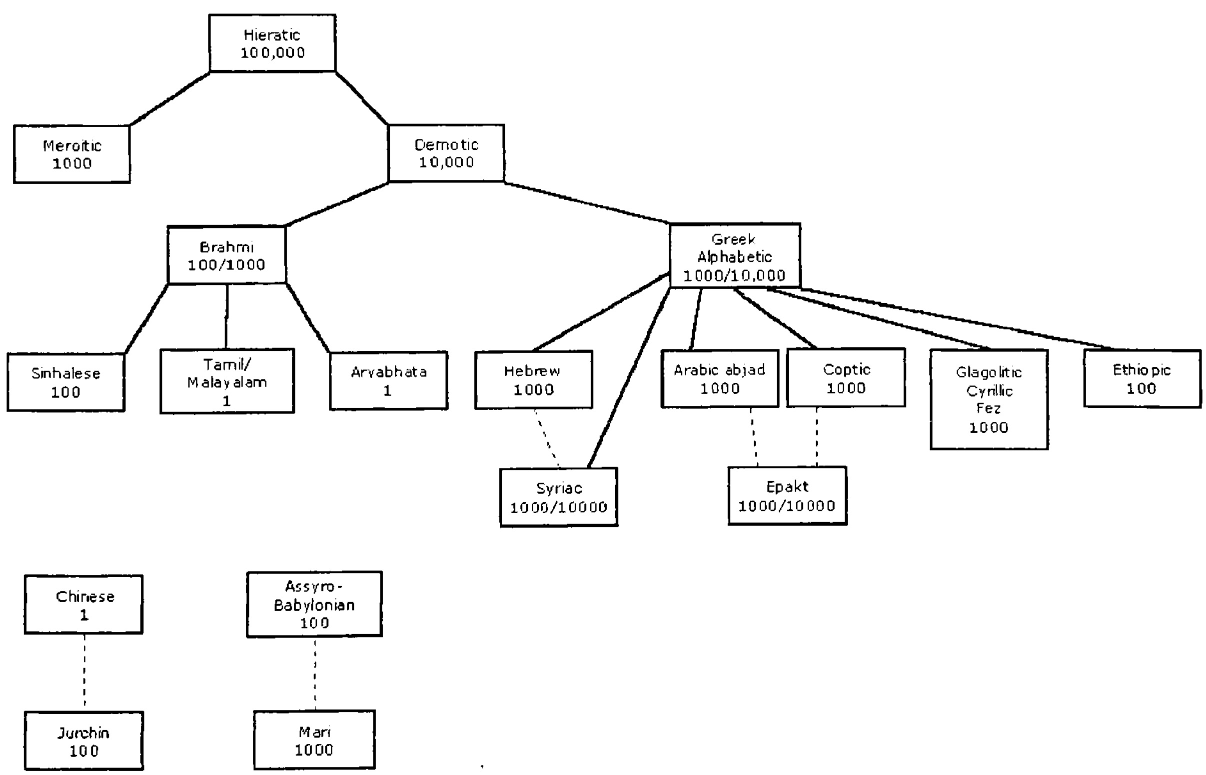


Figure 11.2: Changes in hybrid multiplicative exponent

A final type of transformation involves changes relating to bases and sub-bases. These changes, which are relatively rare, are summarized in Table 11.3.

Table 11.3: Transformations involving base structure

Ch	Ancestor	Base (Sub)	Descendant	Base (Sub)
4	Linear B	10	Etruscan	10 (5)
6	Brāhmī	10	Âryabhata	100 (10)
7	Sumerian	60 (10)	Assyro-Babylonian & Eblaite	10
9	Maya bar-and-dot	20 (5)	Maya head-glyph	20
9	Maya bar-and-dot	20 (5)	Aztec	20
9	Aztec	20	Texcocan & Kingsborough	20 (5)
10	Western	10	Iñupiaq	20 (5)
10	Western	10	Obɛri Ɔkaimɛ	20

No regularities emerge relating to this factor, and in any case, there are too few instances of changes involving bases for any patterns to be discernable. In individual cases, sub-bases may be adopted or abandoned with various cognitive consequences as discussed above, but no diachronic pattern exists. Changes in base are much less frequent than changes in principle. This is probably due to the overwhelming prevalence of decimal lexical numerals in languages worldwide; there is little reason to adopt a new base when developing a numerical notation system unless one's lexical numerals differ in base from that of the ancestral numerical notation system. Curiously, though, four of the changes in base listed above (Brāhmī → Âryabhata and the three Mesoamerican changes) do not correspond to any change in lexical numerals.

Replacement of Systems

The second process governing diachronic patterns concerns the extinction of systems and their replacement by other systems, regardless of any phylogenetic relation between the two. In this section, I use the term 'replacement' to refer only to systems that are totally replaced by other systems, rather than ones that may be replaced only for a limited set of functions, while continuing to be used regularly for others. Doing so will of necessity obscure cases where a system is retained for only limited functions. The replacement of systems is far more frequent than the transformation of systems, because one system may replace many systems, but rarely does one system give rise to multiple systems that use a different principle (cf. Table 11.1). Moreover, a system may be replaced by one that has the same basic structure, which (according to my definition) is not possible for the transformation of systems. Table 11.4 compares the structures of extinct systems with those of the systems that replace them (including cases where the supplanted system has the same structure as the displaced one). As in Table 11.2, I first compare replacement by system (considering both dimensions of the basic structure) and then by each individual dimension.

Table 11.4: Replacement of systems (by principle)

Extinct System	Replaced by					
	Cu-Ad	Cu-Po	Ci-Ad	Ci-Po	Mu-Ad	Total
Cu-Ad	11	0	15	8	1	35
Cu-Po	1	0	1	4	0	6
Ci-Ad	1	0	3	17	1	22
Ci-Po	0	0	0	12	0	12
Mu-Ad	0	0	1	6	2	9
Total	13	0	20	47	4	84

	Intraexponential Replacement		
	Cu	Ci	Mu
Cu	12	28	1
Ci	1	32	1
Mu	0	7	2

	Interexponential Replacement	
	Ad	Po
Ad	35	31
Po	2	16

Two clear trends emerge from the examination of patterns of replacement:

R1. No positional system is replaced by an additive system. There are only two partial exceptions to this rule, both of which involve the replacement of cumulative-positional systems. The quasi-positional system used in the Mesopotamian city-state of Mari in the 18th century BC (ch. 7), which was occasionally used in place of the Assyro-Babylonian system, was eventually replaced by that system after the conquest of Mari by the Babylonians. Yet, as I discussed in that section, positional numeral-phrases were used only rarely and only in the hundreds position; this system is actually best regarded as a short-lived experimental combination of the Babylonian positional (mathematical) and additive (scribal) systems. The second exception involves the replacement of the Babylonian cumulative-positional numerals used in mathematics and astronomy by the Greek alphabetic numerals following Alexander the Great's conquest of Mesopotamia and the gradual domination of Greek over Mesopotamian learning in the exact sciences. Again, this is only a partial exception, because the Greeks borrowed and adopted Babylonian sexagesimal positional numerals in their own mathematics and astronomy, producing the sexagesimal ciphered-positional numerals. This system, though it was only used for fractions, fulfilled many of the functions that the Babylonian system had done, although the role of the ordinary ciphered-additive Greek alphabetic numerals was also important. All other positional systems were replaced by other positional systems (in fact, by ciphered-positional systems) or survive to the present day.

R2. No non-cumulative system is replaced by a cumulative system. There is one exception to this rule. The Gothic numerals of Wulfila's script (ch. 5) were ciphered-additive and were based on the Greek alphabetic numerals. Yet, because they were used primarily in Western and Central Europe, they were replaced by the cumulative-additive Roman numerals. Gothic numerals were used in only a limited number of texts, and so, while other alphabetic systems such as the Greek alphabetic numerals survived and thrived in competition with Roman numerals, the Gothic numerals were overwhelmed. It is surprising that, despite the importance of Roman numerals as an instrument of Roman imperialism, they never totally displaced any of the ciphered-additive systems of Eastern Europe or the Middle East, although they were successful in displacing other cumulative-additive systems, such as the Etruscan numerals (ch. 4) and various cumulative-additive systems of the Levant (ch. 3). It is also notable that despite the historical importance of cumulative-positional systems, such as the Babylonian positional numerals and the Chinese rod-numerals, *no* cumulative-positional system ever totally displaced any other system.

With regard to base structure and the replacement of systems, little can be said in terms of diachronic trends due to the overwhelming prevalence of base-10 numerical notation systems worldwide. All of the base-20 systems of Mesoamerica and the base-60 systems of Mesopotamia were eventually replaced by decimal systems. In only one case was a system replaced by a system with a higher primary base, that being the proto-Elamite decimal numerals (ch. 7), which were replaced by the base-60 Sumerian numerals. It is probably going too far to call this finding a regularity, since so few systems are non-decimal.

The comparison of patterns of replacement with patterns of transformation is instructive, as the effects of the two processes overlap. With regard to interexponential structuring (addition vs. position), positional systems are rarely ancestral to additive systems (except in modern colonial contexts) and tend to replace additive systems over time (but not vice versa). The obvious effect is that positional systems become more frequent over time while additive ones become less frequent. A similar effect is seen in the intraexponential dimension, where non-cumulative systems are rarely ancestral to cumulative ones, and tend to replace cumulative systems over time. Again, the effect over long time periods is to decrease the frequency of cumulative systems. Yet there are also considerable differences between patterns of invention and patterns of replacement. While the intraexponential transformation of cumulative systems into non-cumulative ones is comparatively rare, the intraexponential *replacement* of cumulative systems by ciphered or multiplicative ones is very frequent.

Cognitive Explanations of Diachronic Regularities

In explaining synchronic regularities in terms of the structures of different systems, it was necessary only to show that the presence of a given feature was correlated with some cognitive factor that, if absent, would be inconvenient to the system's users.

Because these regularities were universal or near-universal, these explanations largely involved considerations of hypothetical exceptions. The universality of these rules also means that social context is largely irrelevant to explanations of their existence. In explaining diachronic regularities, we are considering variability among systems, albeit patterned variability. Because of this, we must also ask whether a descendant system is more or less convenient than its ancestor, or whether a successor is more or less convenient than the system it supersedes, with respect to a number of cognitive criteria.

To explain diachronic patterns using cognitive factors, I will compare the observed trends with the advantages or disadvantages of particular features of systems. Where a trend corresponds with increased efficiency in some dimension, there exists a potential explanation for that trend, and thus theories can be developed about which factors were more important to individuals when developing or transforming numerical notation systems. The assumption that one form of representation is advantageous and another disadvantageous can be derived from abstract principles of economy in some cases (e.g. a short numeral-phrase is more advantageous than a long one) or from principles derived from cognitive psychology. As with explanations of synchronic regularities, this is an indirect means of reconstructing cognitive processes, made necessary by the limitations of the data.

It should be expected that explanations of diachronic regularities will be more complex than those of synchronic regularities simply because they must explain change rather than stasis. Cognitive explanations for diachronic regularities explain not only why a feature came into existence, but also how one system is more or less advantageous than another in some respect. This also raises the possibility that advantages as well as disadvantages may be involved in the choice between any particular pair of systems.

Conciseness

The conciseness of a numeral-phrase is simply its length, or the number of signs needed to write that particular number. Strictly speaking, it is thus a property of a numeral-phrase, not of a numerical notation system. All other things being equal, a system that requires many signs to write a number is more cumbersome than one that requires few signs. Because many systems are infinitely extendable, it is impossible to state exactly the average number of signs needed to express numbers, and in any case, this would not necessarily be useful since very high numbers are quite rare. Yet, because a system that regularly requires long numeral-phrases is going to be quite cumbersome to use, we wish to know in general whether a system's numeral-phrases are long or short in order to evaluate its representational efficiency. I will use as a rough measure of a system's conciseness the average length of its numeral-phrases for all numbers from 1 to 999. The principles of numerical notation systems, ordered from most to least concise, are as follows: ciphered-additive → ciphered-positional → multiplicative-additive → cumulative-additive → cumulative-positional. Table 11.5 shows the conciseness of each principle (presuming a base-10 system with no sub-base for each case) for a variety of numbers.

Table 11.5: Conciseness of systems

	Number of signs required (base-10 system, no sub-base, no hybrid multiplication)				
	Ciphered-additive	**Ciphered-positional**	**Multiplicative-additive**[8]	**Cumulative-additive**	**Cumulative-positional**
Number	(Georgian alphabetic)	(Western)	(Chinese classical)	(Egyptian hieroglyphic)	(Inka)
6	1	1	1	6	6
27	2	2	3	9	9
70	1	2	2	7	8
100	1	3	1	1	3
400	1	3	2	4	6
649	3	3	5	19	19

[8] I am presuming here a system like the Chinese classical system (ch. 8) and most other multiplicative-additive systems, but *not* like the Kingsborough Codex numerals (ch. 9), where the unit-sign is in reality a cumulative numeral-phrase.

870	2	3	4	15	16
2003	2	4	3	5	7
4268	4	4	7	20	20
9080	2	4	4	17	19
Average (1-999)	2.70	2.89	4.49	13.59	13.78

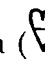
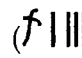
In general, ciphered systems are the most concise, requiring only one sign per exponent. All other factors being equal, no system, for *any* natural number, is ever more concise than a purely ciphered-additive system. While ciphered-positional systems are usually more concise than non-ciphered ones, for round numbers they are sometimes *less* concise because they require zero-signs in the empty positions (e.g. Roman numeral C = 100). Nevertheless, cumulative systems are almost always less concise than their ciphered and multiplicative counterparts, even for low and/or round numbers. Multiplicative-additive systems are slightly less concise than ciphered systems, because they often require two signs (a unit-sign and an exponent-sign) where the latter need only one. Yet, because they are additive, they do not require a zero-sign and are thus more concise than ciphered-positional systems for round and nearly-round numbers. Additive systems are only slightly more concise than positional systems that use the same intraexponential principle; the difference in their conciseness is equal to the number of empty or zero positions required in the number, for which positional systems need a zero-sign. The comparative effect of this difference is not nearly as great as that between cumulative and non-cumulative systems.

The use of bases higher than 10 has variable effects on a system's conciseness, depending on which principle the system uses. Cumulative systems become far less concise through the use of higher bases; in a pure base-20 cumulative-additive system, such as the Aztec numerals, each sign may be repeated up to 19 times, so that 399 requires 38 signs instead of only 20 in a base-10 system (average 22.82 signs/numeral-phrase from 1-999). Yet, for a non-cumulative system, using a higher base makes numeral-phrases slightly *more* concise! In a ciphered-positional system like the Obɛri

Ɔkaimɛ numerals (ch. 10), all numbers from 1 to 19 require only a single sign each, from 20 to 399 only two signs, and from 400 to 8000 only three signs (average 2.58 signs/numeral-phrase from 1-999). These effects have consequences on the sign-counts of these systems (for which see below) that mitigate their positive or negative qualities.

The use of sub-bases is common in cumulative systems as a means of increasing conciseness. By introducing a sub-base of 5 into a decimal cumulative-additive system, a number such as 870 requires only 6 signs instead of 15 (DCCCLXX vs. CCCCCCCCXXXXXXX), and over all numbers less than 1000 the average conciseness is reduced from 13.59 signs per numeral-phrase to 7.45 signs per numeral-phrase. While this reduces the disadvantage of cumulative systems over non-cumulative ones, it never eliminates it. A cumulative system with a sub-base has additional round numbers, which are often expressed as or more concisely than in other systems. Whereas only 14 numbers less than 1000 are expressed as or more concisely in a cumulative-additive than in a ciphered-positional system, the introduction of a sub-base of 5 into the cumulative-additive system raises that number to 54. It is likely that the absence of sub-bases in most non-cumulative systems, and their relative frequency in cumulative ones, is due to the enormous advantage in conciseness that a sub-base provides to the latter but not to the former.

While the use of subtractive notation will be familiar to most readers from the Roman numerals, it is rare in numerical notation systems, being found only in Roman numerals and some of the cuneiform systems of Mesopotamia. It does carry a considerable advantage in conciseness, since 1999 in additive Roman numerals is MDCCCCLXXXXVIIII but MCMXCIX (or even MIM) when subtractive notation is used. However, because subtractive numeral-phrases do not group similar signs together, parsing and reading them may be more difficult. The relative paucity of subtractive numeral-phrases in Roman numerals prior to the modern era, coupled with the fact that subtractive notation is cross-culturally rare, suggests that its advantages were not perceived as being great.

The use of a hybrid multiplicative component for higher exponents of some cumulative-additive systems and ciphered-additive systems has a slightly negative effect on the conciseness of systems that possess this feature vis-à-vis similar systems that lack it. For example, a ciphered-additive system that has no multiplicative component, such as the Georgian alphabetic numerals (ch. 5), expresses 4000 with one sign () while a similar system with a multiplicative component for higher exponents, such as Sinhalese (ch. 6), requires two signs (). Similarly, where a purely cumulative-additive system, such as the Greek acrophonic numerals, requires four signs to write 4000 (XXXX), a hybrid multiplicative system, like the Phoenician system, requires five (). The advantage of using multiplicative expressions for higher exponents is that it obviates the need to develop distinct signs for each multiple of each higher exponent. Presumably, this offsets the slight disadvantage in conciseness.

In terms of the trends observed above, the preference for non-cumulative systems over cumulative ones strongly accords with their far greater conciseness. On the other hand, the trend in favour of positional systems over additive ones does not have a basis in conciseness, since additive systems are slightly more concise than their positional counterparts.

Sign-count

This criterion for evaluating systems is simply the total number of signs a user must know in order to read and write numbers. A system with a smaller sign-count is generally easier to learn and use than one with a larger sign-count because of the decreased mnemonic effort involved. For systems such as the Western numerals, the sign-count is obviously 10. Yet this seemingly simple definition gives rise to nearly insurmountable complexities when attempting to enumerate how many distinct signs a user of a system requires. For systems such as the Phoenician numerals (ch. 3) and the republican Roman numerals (ch. 4), there are multiple signs for many numbers, some of which represent regional or diachronic variability, while others may be multiple signs that every user needed to learn. In other cases, certain signs (normally for very high numbers) developed late in a system's history, were only used by a very few writers, or are non-standard in some other way. In other cases, where a sign is composed of two largely undisguised other signs, a rather arbitrary decision must be made whether to count it as a separate sign. Should the Sumerian sign for 36,000, , be counted as a sign separately from its constituent parts, (3600) and < (10)? Finally, and perhaps most importantly, the issue of sign-count cannot be considered properly without also considering the numerical limit of a given set of signs; the Indus numerals may only have two signs, but these can only express numbers from 1 to 99, whereas the Western numerals have 10 signs but can express any number.

Despite these reservations, it is possible to compare the sign-count of systems based on different principles, presuming that all other factors are equal. The sign-counts of these systems, from lowest to highest, are as follows: cumulative-positional → cumulative-additive → ciphered-positional → multiplicative-additive → ciphered-additive. Cumulative systems, which rely on the repetition of a small number of identical signs, are far less concise but far more economical in sign-count than ciphered ones, which use a wider variety of unrepeated signs. Positional systems, which do not require additional signs to be invented for successive exponents, are more economical in sign-count than additive ones, although there is the additional need to introduce a sign for zero. Cumulative-positional systems, which combine both of these advantages, have extremely small sign-counts (one to three distinct signs). The sign-counts of cumulative-additive systems are very low, but are also dependent on their extendability; a decimal cumulative-additive system without a sub-base requires only one sign for each exponent of 10 that can be expressed (usually four to seven signs, with more if a sub-base is used). Multiplicative-additive systems have slightly larger sign-counts than ciphered-positional ones because, while a ciphered-positional system needs only signs for 1 up to the system's base and 0, multiplicative-additive ones need separate signs for each exponent of the base. The sign-count for a ciphered-positional system is normally equal to its base, while that of a multiplicative-additive system equals its base plus one sign per exponent expressed. Ciphered-additive systems, which require one sign for each multiple of each exponent of the base, have extremely large sign-counts, normally 20 or more, although the cognitive cumbersomeness this would entail is sometimes reduced by the use of script-signs as numeral-signs. Table 11.6 lists some systems whose sign-counts are relatively unambiguous, thus allowing them to be compared.

Table 11.6: Sign-counts (selected systems)

System	Ch	Structure	Sign Inventory	Sign-Count
Western	6	Ciphered-positional	1234567890	10
Old Babylonian positional	7	Cumulative-positional	𒁹 𒌋	2
Egyptian hieroglyphic	2	Cumulative-additive	𓏤 𓎆 𓍢 𓆼 𓂭 𓆐 𓁨	7
Tamil	6	Multiplicative-additive	௧ ௨ ௩ ௪ ௫ ௬ ௭ ௮ ௯ ௰ ௱ ௲	12
Gothic alphabetic	5	Ciphered-additive	𐌰 𐌱 𐌲 𐌳 𐌴 𐌵 𐌶 𐌷 𐌸 𐌹 𐌺 𐌻 𐌼 𐌽 𐌾 𐌿 𐍀 𐍁 𐍂 𐍃 𐍄 𐍅 𐍆 𐍇 𐍈 𐍉 𐍊	27

The effects of base structure on sign-count are variable. For cumulative systems, it does not matter whether a system's primary base is 10, 20, or 60, since a single sign will suffice, as long as it can be repeated as often as is necessary. For non-cumulative systems, however, using a higher base than 10 is extremely detrimental, requiring many more signs to be developed. The only non-cumulative systems with bases higher than 10 are the Maya head-glyphs (ch. 9; base-20, but which use a sub-base of 10 to decrease mnemonic effort), the Obɛri Ɔkaimɛ numerals (ch. 10; base-20, used only briefly and by few individuals), and Âryabhata's numerals (ch. 10; base-100, uses script-signs to decrease effort *and* has a sub-base of 10).

The use of a sub-base also has variable effects on a system's sign-count, depending on the system's structure. For a cumulative system, introducing a sub-base increases its sign-count slightly. A cumulative-positional system requires only one extra sign (for the sub-base), while a cumulative-additive system requires one extra sign per exponent (compare the Egyptian hieroglyphic numerals with the Roman numerals, for instance). In either case, this increase in sign-count is offset by an enormous saving in conciseness; it is safe to say that a base-60 cumulative-additive system, such as the Sumerian cuneiform numerals (ch. 7), could not exist without a sub-base. Yet, in the one ciphered system that has a sub-base (see rule 13), the Maya head-glyphs (ch. 9), introducing a sub-base actually *decreases* the sign-count; instead of signs for 0 through 19, it only requires 14 signs (for 0 through 13) with 14-19 (and sometimes also 13) written with glyphs combining 10 with 4 through 9.

Finally, the use of hybrid multiplication greatly reduces the sign-count of ciphered systems, but has minimal benefit on cumulative systems. One of the major advantages of hybrid multiplication is that a single multiplicative exponent-sign can be combined with a set of existing ciphered unit-signs (1-9 in a decimal system) to avoid needing new signs for each multiple of each exponent. Thus, we see that most ciphered-additive systems of the Hieroglyphic (ch. 2), Alphabetic (ch. 5), and South Asian (ch. 6) families, as well as systems such as the Jurchin (ch. 8) and Cherokee (ch. 10), use hybrid multiplication to express large numbers, while avoiding the need to add nine new signs for each exponent of the base to their already substantial sign-counts. On the other hand, the use of hybrid multiplication in cumulative systems, such as those of the Levantine family (ch. 3) and many of the later Mesopotamian systems (ch. 7), has no effect on sign-count except in very limited circumstances. Where each successive exponent has its own exponent-sign, and exponent-signs are combined only with signs from 1 up to the base (e.g. Aramaic = 100, = 1000, = 10,000, each of which is combined with up to nine cumulative strokes), there is no economy of sign-count; all that multiplication does is avoid repeating signs *other* than the unit strokes; e.g. instead of for 400). Yet, in other systems, such as the Assyro-Babylonian common system, whole cumulative-additive numeral-phrases, including both unit-signs and signs for higher exponents, are combined multiplicatively with large exponent-signs, so that 10,000 was written as 10 ( ) times 1000 ( ), 100,000 as 100 ( ) times 1000 ( ), and so on, eliminating the need for new signs for higher exponents of 10. In such cases, there is a moderate saving in sign-count.

Other than the reversal of the positions of ciphered-positional and multiplicative-additive systems, there is an inverse correlation between a system's conciseness and its sign-count. Thus, the observed diachronic trend from cumulative to non-cumulative systems does not correspond to the much smaller sign-count of the former. On the other hand, the trend towards positional over additive systems may have such a basis, although ciphered-positional systems such as our own have larger sign-counts than most cumulative-additive systems and only slightly smaller ones than multiplicative-additive systems.

Extendability

A system's extendability is measured by the largest number that can be written with it. Unlike conciseness and sign-count, both of which are relevant to the writing even of low numbers, infinite extendability, which is characteristic of most positional systems, only becomes particularly important once there is a strong functional need in a society to express very large numbers (especially where numerals are used commonly for mathematics). However, any increase in extendability - even the addition of a new exponent-sign to an additive system - can be considered as an increase in the capabilities of a system to represent numbers, regardless of the specific functions for which such developments are used.

Most historians of mathematics who have considered the extendability of systems have considered the question only in the abstract, which allows one to say only that positional systems are infinitely extendable while additive ones are not. Consideration of the data shows the situation to be more complex. I have shown that some multiplicative-additive systems are infinitely extendable (rule N1), while some orientational positional systems are not infinitely extendable (rule N2). Nevertheless, the general rule that positional systems are infinite in scope while additive ones are not is largely correct. Even so, not all additive systems are created equal; some are much more easily extended than others. Among additive systems that are not infinitely extendable, systems that use multiplication, whether throughout the system (fully multiplicative-additive systems) or only for larger exponents (hybrids) are generally more capable of expressing higher numbers than those that do not. This is because in many such systems, exponent-signs may be multiplied by entire numeral-*phrases* rather than single signs and/or because multiple exponent-signs placed side by side can be used to indicate repeated multiplications. Most pure multiplicative-additive systems can express numbers as high as 100,000, and many permit numbers as high as 10 million to be written. In the abstract, there is no reason why ciphered systems should be more extendable than cumulative ones, but in practice, they are slightly more extendable, usually having limits of 10,000 or higher whereas many cumulative-additive systems are only used for numbers up to 1000 or 10,000 (such as the modern Roman numerals).

The use of a particular base does not dictate the numerical limit of a system *per se*, because one must also take into account how many exponents of the base the system can represent. Nevertheless, all other things being equal in terms of sign-count and conciseness, a system with a base higher than 10 can represent larger numbers than a base-10 system (e.g. $20^3 = 8000$, so the Aztec numerals can represent any quantity up to 160,000 using only four different symbols, whereas a similar base-10 system could only represent numbers below 10,000). The use of sub-bases has no effect on extendability.

There is an obvious correlation between the greater extendability of positional systems and the trend over time towards positionality over addition. It should be noted, however, that practically any numerical notation system can be extended without great difficulty should the need arise, either through developing new numeral-signs or introducing a structural change such as hybrid multiplication. Where such changes have not been made, it seems clear that there was no overwhelming need for them (at least, not on a regular basis). A great preponderance of the numbers used in both pre-modern and modern contexts are below 1000, and nearly any numerical notation system can deal with such small quantities. Infinite extendability is really only relevant in mathematical contexts.

Effect of Cognitive Factors

To evaluate the overall relation of conciseness, sign-count, and extendability to the diachronic patterns observed above, it is necessary to see how they interact with one another in systems of a given structure. Table 11.7 summarizes these effects on systems of each of the five basic combinations of principle (presuming all other factors to be identical). Each principle is ranked on the three criteria I have discussed (1 being best, 5 being worst).

Table 11.7: Ranking of systems by cognitive factors

	Conciseness	Sign-count	Extendability
Ciphered-additive	1	5	4
	2.70	18-30	Normally 10K – 1 million
Ciphered-positional	2	3	1
	2.89	10-11	Normally infinite
Multiplicative-additive	3	4	3
	4.49	12-14	Normally 100,000 +
Cumulative-additive	4	2	5
	13.59	4-7	Normally 1000 – 100,000
Cumulative-positional	5	1	1
	13.78	1-3	Normally infinite

The most obvious result from this table is that the conciseness and sign-count of a system of a given principle are inversely correlated, except that ciphered-positional systems have a slightly smaller sign-count and are also slightly more concise than multiplicative-additive systems. This correlation is not a coincidence, of course, because one of the most effective ways to increase conciseness is to reduce many signs (one-to-one correspondence) to one, which must involve inventing new signs. Yet there is no correlation between conciseness and extendability or between sign-count and extendability. Systems that are very concise may be highly extendable (ciphered-positional) or limited (ciphered-additive), just as systems with small sign-counts may be highly (cumulative-positional) or less (cumulative-additive) extendable. The reason for this is that conciseness, sign-count, and extendability are properties of *dimensions* of systems (intraexponential or interexponential), not of the systems themselves: ciphered systems are the most concise, cumulative systems have the smallest sign-counts, and positional systems are the most extendable. Because each system is structured both intra- and interexponentially, any system, regardless of the principles it uses, will be less than optimal in at least one of these dimensions. Our own vaunted ciphered-positional system is less concise than the Greek alphabetic numerals and has more signs than the cumulative-additive Roman numerals, both of which it has replaced over time. It is a good compromise between maximum conciseness and minimum sign-count, but it is maximally efficient in neither respect. Moreover, if we assume that cognitive factors are relevant to the invention and adoption of systems (as I think we must), then we may hypothesize about the factor(s) that were most influential in the decisions made at the time of the invention or replacement of specific systems. Of course, we would ideally like to test empirically whether the relevant factors were in fact operant, but this is not normally possible.

A major problem arises when we attempt to extend the analysis of cognitive-structural motivations of specific instances of invention or replacement to produce general rules. The diachronic trend towards positionality over addition, and towards ciphering and multiplication over cumulation, suggests that addition and cumulation should be seen as negative or inferior principles - and of course, this is just the sort of statement that many historians of mathematics have made. To do so neglects an important consideration, which is that cumulative systems are more common than ciphered ones and additive systems are more common than positional ones, and that for many millennia cumulative-additive systems were the most common type. Moreover, if we explain the trend towards ciphering as a desire to maximize conciseness, we must deal with the fact that the trend towards positionality is in opposition to this desire, since positional systems are less concise than their additive counterparts.

To explain long-term diachronic trends, we must acknowledge that the weighting of the cognitive advantages and disadvantages of different principles was not equal in all time periods or in all social contexts. Where there are diachronic trends - such as that favouring non-cumulative systems over cumulative ones - it is entirely likely that they result from changing evaluations of the importance of various merits and defects of different principles. If we want to understand why those evaluations might have changed - for instance, why ciphering (and thus conciseness) came over time on a worldwide basis to be preferred over the small sign-counts of cumulative systems - we must understand the historical conditions under which such evaluations were made. To do so requires that we go further in the analysis of systems than simply looking at their structure, and examine how they are actually used.

The question of diachronic trends becomes even trickier when we examine the overall effects of features other than intraexponential and interexponential principle that I have already discussed in each section above. Table 11.8 summarizes these effects.

Table 11.8: Overall effects of other features

	Conciseness		Sign-count		Extendability
	Cumulative	**Other**	**Cumulative**	**Other**	
Base > 10	Much less	Much more	No effect	Much higher	Higher
Sub-base	More	N/A	Higher	N/A	No effect
Hybrid multiplication	Slightly less	Slightly less	Usually none	Much lower	Higher

The presence of any of these features may mitigate any negative effects or reduce the advantages of the use of a given principle. Moreover, their effects on conciseness and sign-count vary depending on a system's intraexponential principle, adding an additional layer of complexity to the analysis of its merits and disadvantages. I have been unable to determine any diachronic trends relating to bases and sub-bases; hence the effects of such features, if any, on the overall pattern of transformation and replacement of systems seems to be minimal. In fact, although we can speak of the cognitive merits and disadvantages of a system's base and/or sub-base, because these are often a consequence of the lexical numerals of its users' language(s) rather than the result of a conscious decision to alter a system, we would not expect diachronic trends to exist for this feature. Hybrid multiplication is often a flexible way of gaining certain advantages (greater extendability at little extra cost in sign-count or conciseness). Yet the only diachronic regularity concerning hybrid multiplication, rule T4, governs the *degree* of hybrid multiplication (the point above which multiplication is used), not its simple presence or absence – a factor whose cognitive effects I have not analysed because its effects are extremely complex and depend on other structural features of the system.

While the efficiency of systems is obviously relevant to the diachronic patterns I have described, it is not always the case that any potential improvement in a system relating to these factors will be perceived automatically and regarded as relevant by its users. There exist levels of difference too small to be relevant to users, and perhaps too small to be perceived by users. For instance, the minimal difference in conciseness between ciphered-additive and ciphered-positional systems, while recognizable, does not appear to exceed a minimum threshold level (above which, presumably, the additive would be preferred over the positional), while more salient features such as the much smaller sign-count of ciphered-positional systems are probably quite relevant. Any change in a system that would result in ambiguous or poorly ordered numeral-phrases will not register as useful, even if such a change would bear some other benefit. Where significant social factors, such as political hegemony, are involved in the transformation and replacement of systems, otherwise important considerations of efficiency may be irrelevant to users of numerical notation.

In summary, there is no single goal to be attained or variable to be maximized in numerical notation. Every principle has advantages and disadvantages, the choice of which is no doubt governed at least in part by considerations of those qualities, but explaining the diachronic trends observed from the data requires that we ask why certain qualities would be preferred over others. Because four of the five basic principles – the exception, ironically, being the 'ideal' ciphered-positional system - have been developed independently multiple times, we may presume that these systems are perceived as being advantageous, and we can identify the advantages that lead to their adoption. We are thus faced with a situation where the changing functions for which systems are used will be an extremely important factor determining which features of systems will be valued most highly. Thus, any solely cognitive explanation of diachronic regularities will be incomplete.

Summary

Because both the synchronic and diachronic regularities I have outlined relate to various features of systems, cognitive factors relating to systems' structures must be involved in explanations of those regularities. Yet, because the unit of analysis for the two types of regularities is different, the types of explanations involved are quite distinct. A strong case can be made for the parsimonious explanation of many synchronic regularities by cognitive factors alone, since these regularities apply regardless of social context or the specific functions for which systems are used. Yet, even where this is so, we must be careful not to assume that synchronic regularities are consciously imposed rules for the construction of systems; rather, they are outcomes of cognitive processes that arose in specific social contexts (many of which are now lost to us forever). Furthermore, these rules do not do anything to explain the variability among systems, which is still considerable despite the existence of many constraints.

The existence of diachronic regularities is one way to begin to explain the variability that exists among systems. To ignore structural and cognitive features entirely would be ridiculous, given that the trends that exist express tendencies towards particular sorts of systems (specifically, intraexponential ciphering and interexponential positionality). However, there is no perfect numerical notation system; all systems have advantages and disadvantages. To assume that every feature of a system is relevant to its retention or replacement, or that any difference in structure between two systems must have been perceived as important, is erroneous. In order to explain diachronic regularities fully, we must turn, then, to the question of how systems are used, and how systems' functions change alongside the systems themselves, which can only be answered by a careful comparison of specific situations in the history of numerical notation.

Chapter 12: Social and Historical Analysis

Many scholars who have studied numeration have taken great pains to deny that social, political, and ideological factors have significantly influenced the invention, transmission, adoption, modification, and extinction of numerical notation systems. According to their models of invention and diffusion, only the cognitive efficiency and utility of differently structured systems are relevant for understanding the overall history of numerical notation. The myth that the Western numerals achieved their present supremacy solely because of their overwhelming functional superiority over other systems has persisted in this manner. It is easy to see the basis on which this assumption has been made. There are solid reasons for believing that numerical notation is less affected by ideological factors than, for example, the concept of divine kingship (Trigger 2003: 71-91). Numerical notation is a communication system, the primary function of which is to communicate numerical values. One cannot even lie effectively about how many enemies were killed in battle if the numerals being used are incomprehensible to the intended audience. The synchronic regularities I outlined in Chapter 11 provide strong evidence of the limits imposed on the structure of numerical notation systems, without which any system would be essentially unusable by human beings. Because these features are universals and near-universals, social explanations cannot be very useful for explaining them. Even diachronic regularities, which presume the existence of changes in systems, show strong trends towards certain principles and away from others.

Nevertheless, I cannot sustain the strong functional hypothesis that considerations of efficiency are the sole or even the primary influence on the cultural evolution of numerical notation. While synchronic regularities may be explainable without reference to social context, diachronic regularities are not. As I have already shown, every cognitive advantage associated with a system is associated with certain disadvantages. The role of various social factors in explaining the history and development of numerical notation systems will differ from case to case, depending on historical context, but they are always there. No scholar has ever attempted to explain the replacement of Maya numerals by Western ones with reference only to the systems themselves and without consideration of the enormous social, political, and technological upheavals that were associated with the Spanish conquest of Mesoamerica. Nevertheless, in many other instances where the role of social factors is less obvious, many scholars interested in numeration seem willing to ignore the messy complexities of history and rely solely on cognitive criteria. Yet numerical notation systems never exist as objects in isolation; their utility is not merely a function of their structure.

By exploring the social contexts in which the transformation and replacement of numerical notation systems occur, it will be possible to evaluate the impact of social factors relative to purely cognitive and structural ones. I have identified 17 factors that influenced the changes in numerical notation systems examined throughout this study, any of which may be involved in explaining a particular historical event. I list them below, roughly in the order of their importance. These factors are never entirely absent from any instance of systemic change. Historical changes in numerical notation systems can never be explained solely by reference to the structural properties of those systems. The events I will seek to explain are not only the transformations and replacements of systems that I discussed in Chapter 11, but also the geographical diffusion of systems into new regions without structural change and regardless of whether any existing systems were replaced. While I believe that some of these factors may be more important than others (and some clearly occur more *frequently* than others), I do not think it useful to weigh their various effects on the history of numerical notation systematically in the way that I evaluated the differential effects of cognitive factors in the previous chapter. Instead, by providing relevant examples and showing ways that these factors relate to one another *and* to the cognitive factors I have already discussed, I hope to show that there are considerable complexities in the history of numerical notation that cannot simply be reduced to one or a few prime movers. Some of these 17 factors have directly opposite effects to others. There is no contradiction implied in this; rather, it is to be expected, given that multiple goals may be pursued by users and that such multiple interests may need to be reconciled in any given social situation.

Social Dimensions of Numerical Notation

1. A system may be transformed or replaced because its structural features are disadvantageous for new functions for which numerical notation is required.

As I discussed in the previous chapter, numerical notation systems possess inherent efficiency-related characteristics, such as conciseness, sign-count, and extendability. Yet, while these characteristics exist independently of the social contexts in which systems are used, the evaluation of the efficiency of systems requires that we consider which characteristics are most relevant to the particular functions for which a system is used. The analysis of utility must always be linked to the analysis of function. No system is absolutely "efficient" in the way it might be absolutely positional or absolutely decimal. One of the factors we should always consider when a system is transformed or replaced is any possible change in the needs of its users with respect to the writing of numbers.

There are numerous instances throughout this study where systems have changed or been replaced because of the changing social needs of their users. The development of the Babylonian positional numerals (ch. 7) in the late Ur III or early Old Babylonian periods was clearly the result of a new desire to perform mathematics, since the early texts containing experiments with positionality are solutions to arithmetical problems. Similarly, the development of a variant Armenian system by Shirakatsi (ch. 5) was designed to facilitate the mathematical and astronomical work he was doing. The development of Texcocan variants of the Aztec numerical notation system (ch. 9) seems also to have been motivated by new demands relating to land mensuration and surveying in highland Mexico in the early colonial period.

Yet the changes involved need not be so drastic as to produce a system that employs entirely different principles. They may be as simple as the introduction of signs for higher exponents due to increasing administrative needs, such as the invention of signs for successively larger exponents of 10 in late republican Rome, as it grew in size and importance: ↀ for 1000, ↂ for 10,000, ↈ for 100,000. When even this did not suffice, the Romans began using multiplicative notation with a horizontal bar (*vinculum*) for 1000 and an enclosing box for 100,000. When the sign for 100,000 was no longer needed in the early Middle Ages (because of reduced social complexity in Western Europe), it disappeared. Later, under competition from the Western numerals and clearly insufficient for double-entry bookkeeping and mathematics, positional variant Roman numerals were developed. The need to represent higher numbers for administration and mathematics is certainly responsible for the development of multiplicative notation above 100,000 in the Egyptian hieratic numerals (ch. 2) and for the development of various sets of signs for very high exponents of 10 in the Chinese classical numerals (ch. 8). This principle is similar to that suggested by Divale (1999) for the development of higher lexical numerals under conditions of increased need for food storage and preservation.

If a system is being used for purposes for which it is unsuited, this may lead to its replacement for that function, if an obvious alternative is available. Thus, Roman numerals were clearly not conducive to double-entry bookkeeping when it was introduced in medieval Italy and a new system (the early ciphered-positional Western numerals, previously used only by mathematicians) was adopted instead. Similarly, the Arabic abjad numerals (ch. 5) gradually were abandoned and replaced with the Arabic positional numerals (ch. 6), as the exact sciences of the Islamic world became increasingly refined and the administrative needs of the Abbasid caliphate grew. It appears that the same sort of process is currently underway in East Asia, where Western numerals or modified Chinese positional numerals are always used in scientific and technological contexts in place of the multiplicative-additive Chinese system. Yet even where functional considerations play some role in the replacement of systems, we should be careful not to assign them too much importance.

2. A system may be adopted or rejected by individuals or groups because of the number of individuals or groups already using it.

In Chapter 11, the primary considerations discussed in relation to the usefulness of systems were cognitive and structural ones, and I have just described how these factors may relate to the functions for which systems are used. Yet these are not the only considerations relevant to whether a system spreads to new areas. Because numerical notation is a form of communication, the number of users of a given system and the need to communicate with those individuals can be extremely significant. A system that is already used by a large number of individuals may be perceived to be useful by others, regardless of its structure or its usefulness for particular functions.

The prevalence of Roman numerals throughout Western Europe can be explained partly by the Roman Empire's domination of the region, but its spread and continued use also had much to do with the popularity of the system throughout the Middle Ages. Thus, even though other systems known to Europeans had many advantages in comparison with the Roman system, it staved off all its competitors until the 16th century. The adoption of Chinese numerals throughout East Asia was in part a consequence of the advantages associated with adopting a well-known and often-used system. Conversely, systems are particularly vulnerable to extinction when they have few users, especially if there is already a popular system in use in a region. Thus, the failure of the Cherokee numerals to be adopted and the systems of West Africa to achieve widespread popularity is in part a consequence of the fact that they never achieved a critical mass of users. In all these cases, the role played by imperialism is also very important (see factor #3), since popular systems also tend to be those used by large and powerful states. Systems are not accepted or rejected solely according to the number of users they have; the choice to adopt a system may relate to the economic or social advantages of doing so within a system of hegemony or the new system may be imposed externally.

This factor is similar in nature to the 'QWERTY principle', which explains the spread of the sub-optimal QWERTY keyboard as a historical accident which it became very difficult to displace once it had achieved a critical mass of popularity. This inertia is partly due to the difficulties involved in learning a new system and because all the keyboards one is likely to encounter are of the QWERTY form. Similarly, popular computer operating systems may achieve near-ubiquitous (even monopolistic) popularity because of the desire of users to employ popular software packages that they are likely to encounter elsewhere. Because such packages are commonly used, it is rational to continue using them, and their abandonment puts one at a significant disadvantage. Numerical notation systems are not difficult to learn, so the disadvantage of having to learn a new system is not important, but the advantage of being able to communicate with many individuals is significant. Once a system reaches a certain number of users, it becomes much more difficult to displace. I will return to this issue below (see 'Systemic Longevity and Phylogenetic Change').

3. A numerical notation system may be imposed on a society under conditions of political, economic, or cultural domination.

There are many circumstances where the adoption of a numerical notation system is stimulated by direct conquest, encapsulation in a tributary system, or the effects of cultural imperialism. In several cases, a system was introduced into a region that previously had no numerical notation system after its conquest or subjugation by a more powerful polity. In other cases, political or economic domination led to the displacement of a society's existing numerals by another system. This process is extremely important, and I will mention only a few of the many instances where it occurred.

Cases where a system spreads into a region with no previous numerical notation system are very common. The Roman numerals did not come to be used throughout

Western Europe because every society needed such a system, but because the numerals were an administrative tool of the Roman Empire. Similarly, the spread of Egyptian numerals among the early Hebrews was facilitated by the economic domination of Egypt over the Levant around 1000 BC. The most notable example is the prevalence of Western numerals throughout the world, accompanying Western European colonialism and imperial domination in regions that previously had no need for numerical notation. In these cases, there was no or minimal competition with other systems. While in some cases (as in West Africa), indigenous systems may be developed on the model of that of the hegemonic power, these are rarely very successful.

In cases where there was a pre-existing numerical notation system in a region, the effect of domination can be best evaluated when the systems of the dominant and subordinate powers are structurally identical (thereby eliminating differential efficiency for specific functions as an explanation). Thus, the replacement of the Etruscan numerals by Roman ones during the late republican period can only be explained by Rome's rising political and economic fortunes and the decline of those of the Etruscan polities. Similarly, the replacement of the Egyptian demotic system by Greek and later Coptic alphabetic numerals was a consequence of Ptolemaic rule, followed later by Christian missionization. This factor was obviously involved in other situations where the indigenous and successor systems were structurally distinct, as illustrated by the drastic and rapid decline of the indigenous systems of the New World after the Spanish conquest. There is simply no need to compare the relative merits of the systems in such circumstances; for political reasons, there was very little possibility that the Maya, Inka, or Aztec systems would survive for long or replace the systems of their conquerors.

This factor, in combination with #1 (dealing with efficiency for certain functions) raises an interesting issue, though one too large to explore thoroughly in this study. The transformation and replacement of numerical notation systems often depends on social needs relating to administration, bookkeeping, and the exact sciences. These functions are among those that allow large and complex societies to dominate less complex ones. Thus, systems that are well-suited for a set of functions related to the exercise of power will tend to be those that replace the systems of regions that have less well-developed institutions. This provides a potential explanation for why cumulative and additive systems tend to be replaced over time, even though the process by which they are replaced is one of sociopolitical domination. While it is probably going too far to claim that numerical notation is an *instrument* of hegemony, it appears to be an adjunct system that supports hegemonic institutions. Thus, when analyzing the history of numerical notation, it should not be forgotten that it is a useful tool for many tasks relating to the exercise of power.

4. A numerical notation system may be invented in a region upon its being integrated into larger socio-economic networks or by elites in emulation of another society.

This factor is related to the previous one, but deals with circumstances where the model system is not used directly by the adopting society, but instead a new system is invented for local administrative use as the adopting society becomes more complex. In such cases, the functional context surrounding the system's invention is probably administration rather than long-distance trade, since the latter circumstance might make it advantageous simply to adopt one's partner's system wholesale. For instance, the systems of the eastern Mediterranean were developed when those societies (Minoan/Mycenaean, Hittite, Phoenician/Aramaic) increased in social complexity upon entering into regional socioeconomic networks that included Egypt and Mesopotamia. The nature of the long-distance trade that resulted was not such that the adoption of a foreign numerical notation system was particularly advantageous, but the need to control production locally and to extract surpluses made it imperative that some such system should exist. In other cases, the invention of a system may be governed not so much by economics as by the desire of local elites to emulate other states. This appears to be one of the causes behind the development of the Armenian and Georgian numerical notation systems (ch. 5), under the influence of Greek-speaking missionaries during the 4th and 5th centuries AD. It appears that the Brāhmī numerals (ch. 6) may have arisen on the model of the Egyptian demotic system (ch. 2) for a similar reason around the beginnings of the Mauryan Empire.

5. A system may be transformed or replaced if it is incompatible with the computational techniques used in a given society.

Throughout this study, I have downplayed the role of computational efficiency for measuring the usefulness of numerical notation systems, because they are rarely used directly for computation in pre-modern contexts. Yet they are often used indirectly to record the results of computations performed using some other technology. Where the structure of a society's numerical notation systems and computational technologies are consonant (for instance, in base structure or in principle), we can expect that the survival of one system will be correlated with the survival of the other. The continued use of Roman numerals in medieval Europe and of rod-numerals in China is due partly to the utility of the abacus and rod-computation, respectively, for arithmetical calculations. The connection between computational technologies and numerical notation was so strong in these cases that the replacement of the former (by pen-and-paper calculation and the *suan pan* or bead-abacus, respectively) actively contributed to the replacement of the latter (in favour of Western numerals and Chinese positional numerals, respectively). Similarly, one of the factors behind the replacement of the multiple proto-cuneiform systems of the Uruk IV period in Mesopotamia (ch. 7) may have been the abandonment of older metrological systems. In the Early Dynastic period, once those systems were no longer used, the corresponding numerical notation systems declined. Another case that may be a result of computational techniques is the development of the Etruscan numerals (ch. 4) out of tally-marks, which are most congruent with a system that uses a quinary sub-base.

It is important to note that this factor is not all-important. The use of the abacus has not declined significantly in Japan despite the very widespread use of Western numerals there. Despite the consonance between the quinary-decimal systems of the Italic family and the use of the abacus, systems such as the Greek acrophonic numerals were replaced with the ciphered and non-quinary alphabetic numerals, even though the use of the abacus continued. Finally, the use of hexadecimal and binary numbers in electronics is not likely to be a factor in the replacement of our current decimal Western numerals.

6. A system may be used for limited purposes in which it is useful to distinguish one series of numbers from another.

Many societies retain older systems for limited purposes so that the two systems, when used together, help distinguish two types of objects, each of which is enumerated using a different system. Doing this may serve to reduce ambiguity or at least to indicate the function of a numeral-phrase by the system that it uses. For instance, in the modern West, Roman numerals are retained for prefaces to books, volume numbers for multi-book series, certain lists (especially those with sub-categories), and sometimes even in dates (6.vii.2002 instead of 6/7/02). In modern Greece, the same principle governs the occasional use of the alphabetic numerals for numbered lists, even though Western numerals are used in most contexts. Ironically, in ancient Greece, the acrophonic numerals were retained for stichometry as late as the 3rd century AD, even though they had been superseded by the alphabetic numerals centuries earlier. A far more ancient example is the employment of multiple systems in Mesopotamia. From their inception, various proto-cuneiform systems were used to express different types of quantity. While Nissen, Damerow, and Englund (1993) have interpreted this as evidence of the absence of abstract numeration at that time, I believe that it is just as likely to have been a simple functional division based on the employment of several different metrological systems. Similarly, in the second half of the 3rd millennium BC, linear-style Sumerian numerals and the newer cuneiform signs were used in the same texts to indicate different types of object, possibly to avoid confusing the different categories when taking sums. In these cases, whatever reason the older system ceased to be used was outweighed by the value of maintaining a second system for auxiliary purposes.

7. At the time of the diffusion of numerical notation into a region, the principle of the ancestral system may be adopted, but an indigenous set of numeral-signs is developed.

The principle of 'strength in numbers' (#2) suggests that the need to be understood by a wide range of users reinforces the spread of already-popular systems. Yet in many cases in this study, even when the structure of a system is adopted precisely, the numeral-signs adopted are indigenously invented, although this change renders them unreadable to users of other systems. One of the most important reasons for doing this is for the adopters to express a different cultural identity than that held by those who transmitted the system, possibly in the process obscuring the origin of the new system. The clearest examples of this are in systems such as the Kpelle, Bamum, and others of West Africa (ch. 10), where ciphered-positional systems were developed on the basis of Western or Arabic numerals, but indigenous numeral-signs were invented. Similarly, in a case such as the possible development of Linear A numerals from the Egyptian hieroglyphs (ch. 2), it would not have made much sense for the Minoans to adopt the hieroglyphic signs, which were also phonetic signs of the hieroglyphic script. Instead, while the Linear A system is structurally identical to its ancestor, it uses simple abstract signs. This factor is also responsible for the many different systems of the alphabetic family (ch. 5), where each script has a distinct numerical notation system using its own letters as numeral-signs. Since the numeral-signs were also script-signs, and were developed at the same time as each script, it would not have made any sense to retain "foreign" numerals, since the very point of alphabetic numerals is the need to learn only one set of symbols. It is important to distinguish this factor from the paleographic divergence of systems that once were unified, such as the changes seen in the Brāhmī-derived systems of India. In such cases, the divergence of systems occurred well after the time of a system's invention, in response to the separation of regions that were once politically unified or the migration of peoples (cf. factor #14).

8. A descendant system may be structurally distinct from its ancestor because of differences in the lexical numerals associated with them.

Systemic transformations often result from efforts to adapt a system to the structure of the lexical numerals associated with the adopting society, particularly to the base of the new system. In certain modern instances, a system's inventors explicitly stated their intention to fit a numerical notation system to their lexical numerals, as in the Iñupiaq and Obɛri Ɔkaimɛ (ch. 10) systems, which are both vigesimal even though they were derived from the Western numerals. In pre-modern cases, usually we can only infer that such a decision was made by comparing a group's numerical notation system and lexical numerals. Most authors presume that the shift from sexagesimal to decimal numerical notation in Mesopotamia corresponds with the shift in dominance from Sumerian to Semitic speakers (although sexagesimal elements were retained in Assyro-Babylonian systems). In some cases, the additional signs of a system rather than its major features are affected. The use of special signs for 11-19 in the Jurchin numerals (ch. 8) corresponds to the fact that in the Jurchin language, the corresponding lexical numerals are not directly related to the word for 'ten'.

9. In a historical context, when an established system is challenged by a new one, the older system may be defended and the interloper denigrated for cultural or political reasons.

It is virtually inevitable that when a new numerical notation system is introduced into a society, there will be competition between its proponents and its detractors. I have already discussed situations where the new system is imposed through conquest or cultural hegemony (#3). In some cases, active and successful local resistance can prevent or delay the new system from achieving a foothold. The effect of cultural inertia or tradition varies according to the historical context and cannot be predicted based on the relative merits of the competing systems. It is probable that resistance to the Western numerals in China and the preference for the Chinese classical numerals fall into this category. While Western numerals quickly took hold in Japan and Korea, in China, the cultural associations of the classical numerals (together with their correspondence to the lexical numerals) meant that resistance was much more effective than elsewhere. Where strong religious connotations are attached to the use of a particular system (as with the Hebrew alphabetic numerals), it may be almost impossible to displace them, even when the system's users are encapsulated in larger polities. As I mentioned in Chapter 6, one reason why the Malayalam, Tamil, and Sinhalese systems remained non-positional for a long time may be that the new invention was perceived to be associated with Hinduism. Yet, in other cases (the Mesoamerican systems come to mind), a generation or two suffices to eliminate a system, and any resistance is overcome relatively quickly. Often, resistance to the new system takes the form of invention of an entirely new system – witness the creation of the Varang Kshiti and Pahawh Hmong numerals in the twentieth century, or the invention of quasi-positional Roman numerals in reaction against the Western numerals. Such systems have rarely been very successful.

Even where the arguments defending one system against another purport to be concerned with efficiency, the role of tradition in resisting new and/or foreign inventions can be quite important. Such sentiments appear to have been behind the prohibition of Western numerals in Florence in 1299 and similar derogatory statements about their ease of forgery in Western Europe between the 13th and 16th centuries (Struik 1968; Menninger 1969: 426-427). Although much of the discourse decrying the merits of Western numerals in late medieval Europe focused on their potential to be used for illicit purposes, it seems highly probable that other factors were at work. These new numerals were a foreign invention and could be seen as undesirable by xenophobic administrators. They were also associated with merchants and moneylenders, and so class interests may have been relevant.

10. A system may be borrowed or invented for use in a limited sector of society to control the flow of information.

While it is relatively rare for systems to develop or diffuse for obscurantist purposes, sometimes a system is developed primarily to conceal information in a code understood by only a limited number or else to protect information against forgery. The siyaq numerals and Turkish cryptographic numerals (ch. 10) appear to have their origins in the desire of certain categories of individuals to control the flow of information. The Cistercian numerals (ch. 10) also may have occasionally been used cryptographically, particularly in the latter part of their history. The Fez numerals (ch. 5), originally used quite widely, were eventually used only in contracts in order to conceal values and thus prevent forgery and modification. A similar function is served by the *da xie shu mu zi* accounting numerals used in China; the complexity of the numeral-signs makes altering these numerals for fraudulent purposes nearly impossible.

11. A system may be retained for prestige or literary purposes even after it has been supplanted by another system.

The retention of Roman numerals in the West is the best-known example of such a situation. They are often used today, in contexts such as clock faces, monumental inscriptions, copyright dates of films, and ordinal numbering (e.g. of monarchs, World Wars, and Super Bowls), to assign prestige value to something by denoting it in Roman instead of Western numerals. They carry with them a connotation of age and classical education, and their retention into the foreseeable future thus seems likely. Similarly, the retention of Greek and other alphabetic numerals, particularly in liturgical contexts, reinforces the venerable status of texts that use them. Particularly elegant forms of the Chinese numerals, such as the *shang fa da zhuan* used on seals, are treasured for their age and beauty. Finally, the retention of Sumerian numerals in certain Assyrian royal inscriptions as late as the 8th century BC, seems to have served the purpose of associating the kings mentioned in those inscriptions with the traditions of ancient Mesopotamia.

12. A system may be invented on the model of two or more existing systems.

In outlining patterns of systemic transformation in Chapter 11, it was sometimes necessary to employ a slight elision in describing the transformational process, by treating all ancestor-descendant relationships as ones with a single ancestor and a single descendant. In most cases, this is accurate; if the inventor(s) of a new system knew and used additional numerical notation systems, these systems had little effect on the structure of the new system and the form of its numeral-signs. In other circumstances, however, a system blends important features of two ancestral systems. This is most noticeable in examining the Levantine family (ch. 3), whose origins lay in the interaction of the hieroglyphic systems of the eastern Mediterranean (most notably the Egyptian hieroglyphs, but also the Hittite hieroglyphic and Linear B numerals) and the Assyro-Babylonian cuneiform system of Mesopotamia. In this case, the two ancestors and their descendant are cumulative-additive, so no change of principle was involved. Yet, to understand fully the development of Phoenician and Aramaic numerals, we must understand how and why these two systems interacted in the way they did. In this instance, Levantine peoples were intermediaries in trade relations between Egypt and Mesopotamia, thus accounting for the fusion of systems from both regions. Another case that does not involve a change in principle is the combination of the Arabic abjad numerals and the Coptic numerals into the cursive 'Epakt numerals' (ch. 5) used in Egypt under the Fatimid caliphate.

A slightly more complex scenario of blending of systems occurs when the basic structure of an existing system is altered because of knowledge of another system. This was the case with the development of Roman positional variants after the introduction of Western numerals into medieval Europe (ch. 4) and the transformation of Âryabhata's numerals into the ciphered-positional katapayâdi system after the Indian ciphered-positional numerals had been invented (ch. 6). In these cases, a structural transformation occurred in which the signs of an older system were combined in a new way on the model of the intraexponential and interexponential structure of another one. In such instances, we may be seeing a pattern of attempted resistance to the innovation by altering the structure of an older system. In other cases, such as the Chinese commercial numerals (ch. 8), which combine the classical system and the rod-numerals, functional considerations, such as the need to do rapid calculations, may have been more important.

13. A system may be transformed or replaced because of changes relating to the media on which or the instruments with which it is written.

The transformation of the Egyptian hieroglyphic numerals into the hieratic system (ch. 2) is the only known instance in history where a cumulative-additive system gave rise directly to a ciphered-additive one. Yet this development was not so much an adjustment to new functions for the system as it was to the new media (ink on papyrus) used in writing it. The adoption of a cursive script tradition and a media on which distinct cumulative signs could be reduced gradually to ligatured ciphered ones was a significant development. Of course, there were functional shifts as well, but the particular nature of the transformation that occurred was largely dictated by the change in medium; it would have been impossible for the hieratic ciphered-additive system to develop within the context of Egyptian monumental writing. Although the Egyptian case is only one example, I suspect that the cursive reduction of many signs to one sign is a far more likely change than the reverse (some sort of division of previously ciphered signs into cumulative ones), and that this explains in part the transformational trend away from cumulative signs.

I suspect that factors relating to writing media may help explain the difference between the ordinary Maya bar and dot numerals used on monumental inscriptions and the rotated and quasi-positional ones used in the Dresden Codex (ch. 9). In this case, however, we have to deal not only with the paucity of surviving Maya texts but also with the contrary fact that some very early Mesoamerican monumental inscriptions are of the quasi-positional type. Changes in writing style in Mesopotamia were responsible for the rotation of signs from top to bottom to left to right in direction, and a later change in stylus shape produced the shift from the Sumerian archaic to cuneiform numerals (ch. 7). The paleographic differences between the ordinary Arabic positional system and the North African *ghubar* numerals (ch. 6) are partly due to the former system's use on stone inscriptions and in texts, whereas the *ghubar* system was used in 'dust-board' calculation.

A quite different instance where this principle applied was in the gradual replacement of Roman numerals by Western ones in early modern Europe. Western numerals can be used with considerably greater ease than Roman numerals in printed books and on dated coins because of their greater conciseness. The adoption of Western numerals may have been influenced to some extent by the increasing use that the burgeoning middle classes of Western Europe were making of books in the 15th and 16th centuries. I do not seek to downplay other factors (e.g. their computational efficiency for mathematics and bookkeeping), but merely to point out another relevant attribute.

14. A system used in multiple politically independent or geographically diverse regions may diverge over time into several systems.

In many cases, a single system diverges over long periods into multiple systems, usually when a previously unified region becomes politically fragmented or because of geographical separation caused by migration. This paleographic drift may or may not cause structural changes but, given enough time, it usually results in systems that are related to one another phylogenetically but are not mutually intelligible. It is quite different from the process by which a system's signs were modified at the time of its adoption (factor #7), although its result is similar. This process was responsible for the initial divergence of the Egyptian demotic and 'abnormal' hieratic numerals that were used in Lower and Upper Egypt, respectively, in the politically fragmented Late Period. Similarly, the fragmentation of Achaemenid Persia after the Alexandrine conquest, and the relatively loose Seleucid rule thereafter, led to the divergence of the older Aramaic system (ch. 3) into its many structurally distinct variants used in the city-states and other small polities of the region: Palmyrene, Nabataean, Hatran, and so on. The best-known example of such a divergence is that by which the Brāhmī numerals used in the Gupta Empire developed into the various Indian systems after the empire's fragmentation in the 6th century. The spread of these numerals into the Arab world and thence to Western Europe continued the process of paleographic divergence so that today it is difficult to see any resemblance among the numeral-signs of the scripts of Europe, the Middle East, and South Asia.

15. A system may diverge structurally from its ancestor due to factors related to the phonetic script of the society.

In a few instances, the structure of a society's script is inconsistent with the way in which a diffused numerical notation system forms numeral-signs, thus forcing or enabling changes in the numerical notation system that eventually develops in the recipient society. For instance, the Greek alphabetic numerals have 24 signs plus 3 episemons, but the Hebrew consonantary has only 22 signs. Thus, when the Hebrew numerical notation system was invented, a new technique had to be invented to represent the numbers 500-900, which was to combine the 22nd sign (for 400) with other signs for 100-400 as necessary. In contrast, in the Greek-derived Armenian and Georgian systems, whose corresponding alphabets had more than 36 signs each, unique signs could be developed for 1000-9000 instead of using hybrid multiplication. The way in which script-signs affect the structure of alphabetic numerical notation systems is discussed extensively by Gamkrelidze (1994). The failure of the various Indian alphasyllabic numerical notation systems (ch. 6) to achieve widespread acceptance is also due in part to this factor. These systems could not have spread to regions that lacked alphasyllabaries because their structure requires that the script with which they are associated be alphasyllabic.

16. An existing system may be retained after its replacement for purposes for which it is more useful than the system replacing it.

This uncommon circumstance is the inverse of #1, and occurs when a system is replaced for most purposes, but retained for a very limited set of functions because the newer system is inadequate in some way. In both the South Asian and Alphabetic families, riddles exist that use the assignment of numerical values to phonetic signs to embed dates or other numerical values in words or phrases. When the Hebrew alphabetic numerals, Arabic abjad, and other systems were replaced by positional numerals for most purposes, the older systems were retained for number-magic because the newer systems did not assign numerical values to letters. Similarly, the *varnasankhya* systems of India (Âryabhata's system, katapayâdi, aksharapallî) were used by astrologers and in literature for centuries after they had been superseded by ciphered-positional numerals, and some such systems are still used today.

17. A systemic transformation may result from factors relating to ideological subsystems of the society in which it is used.

This very rare circumstance nevertheless in two cases resulted in important structural changes. The invention of the ciphered Maya head variant glyphs as alternatives to the cumulative bar and dot numerals seems to have been motivated by the symbolic association of gods with numerical values, probably related to phonetic correspondences between their names and Maya lexical numerals (Macri 1982). The invention of the head variant glyphs therefore appears to have been aesthetically and religiously motivated; the complexity of the glyphs and the consequent difficulty in inscribing them on monuments refutes any simple functional explanation for their development. The second transformation is the development of positionality in India and the shift from ciphered-additive notation to ciphered-positional numerals with a zero. As I discussed in Chapter 6, positionality has some clear literary antecedents in Hindu philosophy of the late Gupta period, including the development of the concept of the concept of *sunya* 'emptiness, void' and the subsequent naming of the zero-sign *sunya-bindu*. While there may also have been functional correlates to this development, this philosophical prefiguring of positionality and zero is nonetheless highly intriguing.

In summary, each of these factors is relevant in multiple systems examined in this study, and no system's development or replacement can be analysed without taking account of the effects of various social circumstances on the historical record. None of these factors refutes the findings of Chapter 11, where I demonstrated powerful multilinear diachronic trends favouring ciphered and positional systems over time. It thus becomes imperative to explain how these social and cognitive factors interacted to produce the attested historical patterns. One important way in which this occurred was through the combination of the increasing functional need of numeration for administration and exact sciences that accompanied the development of social complexity, with the greater potential that such functions allowed for dominating other societies. Once this process had begun, the number of users of such systems increased, which made it more likely that these systems would be perceived as useful by members of other societies. While resistance to introduced systems might be partially successful and might result in the retention of older systems for limited purposes, the diachronic trend seems to have favoured systems whose users are associated with larger-scale and more complex societies.

Systemic Longevity and Phylogenetic Change

I have devoted much attention to the major changes that occurred over the history of numerical notation. These events of transformation and replacement are extremely important from a theoretical perspective, since they help us understand why systems are invented, altered, and replaced. Yet the desire to explain change must be understood in its proper context. Episodes of transformation of numerical notation systems are extremely rare, and the replacement of systems is only slightly less so. Numerical notation has existed for 5,500 years, but there have only been around 20 attested instances of a system giving rise to one that uses a different basic principle (an average of one event every 275 years), and only about 80 instances of a system going extinct (approximately one every 70 years). Thus, in contrast to many other sociocultural phenomena, numerical notation systems are remarkably durable and long-lived. Table 12.1 lists all 28 numerical notation systems that were used for periods of 1,000 years or more, including systems still in use (with BC dates indicated using negative numbers).

Table 12.1: Long-lived systems

System	Principle	First	Last	Duration
Egyptian hieroglyphic	Cu-Ad	-3250	400	3650
Egyptian hieratic	Ci-Ad	-2600	200	2800
Greek alphabetic	Ci-Ad	-575	2000	2575
Roman (classical)	Cu-Ad	-400	2000	2400
Chinese (traditional)	Mu-Ad	-250	2000	2250
Assyro-Babylonian common	Cu-Ad	-2300	-200	2100
Hebrew alphabetic	Ci-Ad	-100	2000	2100
Babylonian positional	Cu-Po	-2000	0	2000
Maya (bar and dot)	Cu-Ad	-400	1600	2000
Chinese rod-numerals	Cu-Po	-300	1600	1900
Ethiopic	Ci-Ad	350	2000	1650
Coptic	Ci-Ad	350	2000	1650
Sinhalese	Ci-Ad	500	2000	1500
Tamil	Mu-Ad	500	2000	1500
Syriac alphabetic	Ci-Ad	500	2000	1500
Indian	Ci-Po	575	2000	1425
Sumerian	Cu-Ad	-2900	-1500	1400
Malayalam	Mu-Ad	500	1850	1350
Maya positional	Cu-Po	-50	1250	1300
Armenian	Ci-Ad	400	1650	1250
Egyptian demotic	Ci-Ad	-750	450	1200
Arabic positional	Ci-Po	800	2000	1200
Georgian	Ci-Ad	450	1600	1150
Arabic (ghubar)	Ci-Po	875	2000	1125
Brāhmī	Ci-Ad	-300	800	1100
Cyrillic	Ci-Ad	900	2000	1100
Shang (oracle-bones)	Mu-Ad	-1300	-250	1050
Siyaq	Ci-Ad	900	1925	1025

This chart demonstrates firstly that many systems thrive for very long periods; the systems on this list comprise nearly one-third of all those examined in this study. Moreover, longevity is not exclusive either to earlier or later systems, as is shown by the presence of both very old systems, such as the Egyptian hieroglyphs, and relatively recent ones, such as the Arabic positional numerals. The duration of some of these systems may be slightly exaggerated due to the fact that many systems survive for several centuries after they fall out of common use: this is certainly the case with the Roman numerals, various alphabetic systems, and the Egyptian systems. Yet the effect of such survival is very small in most cases, and under any calculation the Egyptian hieroglyphic and hieratic numerals have the longest period of use.[1] There is absolutely no correlation between a system's principle and its longevity; all five combinations of principle are found multiple times among long-lived systems. The fact that 13 of the 28 systems are ciphered-additive is an artifact of the larger number of such systems overall. Obviously, there are few long-lived ciphered-positional systems so far, because they are mostly of relatively recent invention and could not yet have existed for 1000 years.

The great longevity of many systems is due to the persistence of several civilizations over long periods. The systems of Egypt and Mesopotamia are among the longest-lived because the cultural traditions of the Egyptian and Mesopotamian civilizations were very stable, there was little impetus to develop new systems, and cultural contact with other regions that might offer alternative systems was limited. Other systems (e.g. the Roman and Chinese classical systems) persisted due both to their use in enormous empires and to their subsequent use as shared numerical notation systems over large regions. Still others were developed and persisted in the context of specific liturgical and literary traditions; this is certainly the case with many of the long-lived alphabetic systems. In all of these cases, change in numerical notation systems is

[1] If we consider the Shang numerals and the Chinese traditional numerals to be part of the same system, however, their total lifespan is 3300 years so far.

the exception rather than the rule. Systems can persist for millennia even in the face of competition from others that may be more efficient for some functions.

Reconciling the many factors that can lead to systemic change with the reality that such changes are comparatively rare requires explanation. At least three separate factors combine to ensure the relative stability of numerical notation systems. The first is that there is no strong selection against systems, even when they are insufficient for the purposes for which they are being used. In most cases, numerical notation systems seem to operate on the principle of the 'survival of the mediocre', which means that a system will tend to persist unless it is obviously maladaptive for the functions for which it is being used (Hallpike 1986: 81-145). Until the rise of mathematics and double-entry bookkeeping, Roman numerals were reasonably well suited to any of the purposes for which they were needed in either classical Rome or medieval Europe. Even where they were perceived to be inefficient, options other than replacement were available. For mathematics, the Greek alphabetic numerals could be used. For mensuration, metrology, and arithmetic, multiplication tables and similar arithmetical charts could be introduced, or other computational techniques such as finger-arithmetic and the abacus could be used. If the Roman numeral-phrases were simply too cumbersome and long, techniques such as subtraction could be introduced. The adoption of a new system is a drastic step. A more adequate alternative system must not only exist, but also be perceived as sufficiently useful to justify abandoning the older system. It is simply not the case that the history of numerical notation can be explained in strongly selectionist terms, whether cognitive or social explanations are invoked to account for the invention and decline of systems.

Secondly, even though it is not difficult for one individual to learn and use a new numerical notation system, the wholesale replacement of an older system throughout a large social network is extremely difficult because numerical notation is used for communication. As such, one of the primary factors governing a system's usefulness is the number of its current users (social factor #2, above). Even if a new numerical notation system introduced into a society has some advantage, it must overcome the disadvantage that initially it has few users and that it is not very effective for communication until some critical mass of them is reached. This is unlikely to occur unless there is a significant shift in social conditions - for instance, the integration of a society into a larger network of trade and interregional interaction, or the imposition of the numerical notation system by force. In such situations, it may be advantageous for the new system to be adopted by certain groups (traders, for instance), by which means it may gradually acquire the critical mass necessary to displace the older system. This circumstance seems to describe quite closely the replacement of the Egyptian systems (ch. 2) by the Greek alphabetic and Roman numerals in the Ptolemaic period and beyond. Even though alphabetic numerals were known and used throughout the region by the 4th century BC, they did not displace the Egyptian systems for many centuries. Greeks and Romans in Egypt employed their own systems, while Egyptian scribes used their indigenous ones, until the number of users of the introduced systems so greatly outnumbered those of the Egyptian ones that there was no other option.

A final factor that may explain the relative stability of numerical notation systems is that they are written rather than verbal. Their stability is directly comparable to the stability of scripts, which also can persist without major change for millennia despite radical social and linguistic changes. The Roman alphabet and Chinese logosyllabary have changed little over the past two millennia, even though the spoken languages associated with them have continued to change radically. Numerical notation systems, like scripts, maintain their stability because older texts are read generation after generation. Retaining an existing representational system ensures that older texts and inscriptions can continue to be read.[2] As long as there is a continuous literary tradition in a region, abandoning an established writing system means that older texts may become confusing and unreadable. Because numerical notation systems are trans-linguistic written systems, an additional factor accounting for their stability is that they may spread very widely, and even if they cease to be used in one region, may be retained in others.

Changes do occur in systems other than episodes of transformation of principle and complete replacement. Paleographic alterations in the shapes of numeral-signs happen regularly in numerical notation systems as they do in scripts, particularly cursive ones. Even in modern Western numerals, there are variant forms for many numeral-signs (0 vs. Ø, 2 vs. 2, 4 vs. 4, 7 vs. 7). These changes, while often inconsequential, sometimes can have great effects (as witnessed by the cursive reduction of Egyptian hieratic numerals from their hieroglyphic ancestors). Even if we are inclined to dismiss paleographic changes as trivial, minor structural changes cannot be dismissed so easily. These include a) changes in non-base numeral-signs that contribute to a system's structure; b) the introduction of subtractive notation; c) the invention of new signs for higher exponents of a system's base; d) changes in the point above which hybrid multiplication is used in a system; e) changes in the direction of writing of a system; and f) changes in the way in which cumulative systems chunk groups of signs.

These minor diachronic changes represent the vast majority of changes that occur in numerical notation systems. Yet it is exactly these minor changes in the structure of systems that distinguish similarly structured systems within each phylogeny discussed in Chapters 2 through 9. In the vast majority of cases, a system uses the same base and the same structural principles as its descendants. Families of systems represent yet further stability in numerical notation, as they are composed of long chains of ancestor-

---

2 This is one of the major reasons why various attempts at Chinese script reform have met with limited success, despite the widespread recognition that the existing script is very difficult to learn.

descendant relationships. Every family in this study has a total lifespan greater than 2000 years, with the exception of the Levantine family which checks in at a 'mere' 1350 years. These are enormous spans of time for recognizable representational systems to survive. Individual systems among these families can fail after only a short time, and several independently invented systems (e.g. Inka, Bambara, Indus) gave rise to no descendants and thus represent abortive phylogenies. Moreover, some families are far more unified than others. While the systems of the Italic family share many structural features, others (such as the East Asian and Mesoamerican families) can be identified as being descended from a common ancestor only through historical context. Even so, the fact that such traditions can be identified by any means highlights the remarkable persistence of numerical notation systems.

Civilization and Systemic Invention

My analysis so far has focused on issues of diffusion, transformation, and replacement of systems rather than on events of independent invention. As I argued in Chapter 1, there is no reason to postulate a qualitative gulf between cases that have a specific antecedent numerical notation system and those that do not. Many of the functional and social needs that govern the adoption of other societies' systems or the invention of new ones using an external model are identical to the ones governing the invention of systems in the absence of such a model. As Julian Steward (1955: 182) maintained, every borrowing must be construed as an independent recurrence of cause and effect. Moreover, because other representational systems (e.g. unstructured or minimally structured tally systems, lexical numerals, metrological systems) precede the independent development of numerical notation, we must recognize that, when we speak of 'independent invention', we are not simply talking about an invention that springs forth from nothing into the mind of its creator. Nevertheless, it is to be expected that the process by which independently invented systems arise may be somewhat different from that by which systems are modelled on a specific ancestor.

In this study, I have identified seven systems that were almost certainly invented independently of any specific influence from other numerical notation systems: the Egyptian hieroglyphic (ch. 2), Mesopotamian proto-cuneiform (ch. 7), Shang Chinese (ch. 8), Maya bar and dot (ch. 9), and the Harappan, Inka *quipu,* and Bambara (all ch. 10) systems. In two additional cases – the Etruscan (ch. 4) and Brāhmī (ch. 6) numerals – the hypothesis of independent invention could not be rejected entirely. In yet two more cases – the Chinese rod-numerals (ch. 8) and the *siyaq* numerals (ch. 10) – while it was clear that their inventors knew other numerical notation systems, these other systems did not play any evident role in their development. Finally, one system – the Aztec numerals (ch. 9) – is historically related to earlier Mesoamerican systems only through the intermediary of the unstructured highland Mexican 'dot-only' system, which was not itself a full-fledged numerical notation system with a base and intra- and interexponential structure.

I find it notable that the development of independently invented numerical notation systems coincides very closely with the rise of civilizations in Egypt, Mesopotamia, East Asia, the Indus Valley, Mesoamerica, the Andes, and elsewhere. Early civilizations are qualitatively distinct from the less complex societies that precede them, being characterized by great socioeconomic inequality, surplus extraction, and a complex administrative apparatus. The state of Teotihuacán, another potential early civilization, used the Mesoamerican bar and dot numerical notation system exceedingly rarely and from available evidence only to express very small numerical values (see ch. 9). Moreover, if the Etruscan and Brāhmī cases are truly independent creations, these also developed in the context of the emergence of civilization in Italy and India, respectively. Yet in the pre-colonial West African Yoruba civilization, there was no numerical notation system.[3] The Bambara system was used further north, but although we know almost nothing about its history, there is no evidence for its use among the Yoruba. There is likewise no evidence of the employment of numerical notation systems in many other African or early Peruvian civilizations. There is thus a very strong correlation between the origin of numerical notation and the emergence of many, but not all, civilizations. This finding suggests that the initial development of numerical notation frequently may be a response to new social needs that arise at a certain level of social complexity. This could also help to account for the development of numerical notation systems in colonial situations.

A difficulty with this proposition is that the functions for which numerical notation was used in these societies are variable. Among the Egyptians, Mesopotamians, Inka, and Aztecs, numerical notation was first used for administrative and accounting functions. Yet in Shang China, the first attested numerical notation is found on oracle-bones and as such seems to have been used for divinatory purposes. From Ganay's (1950) ethnographic work, our best guess is that the Bambara system was also used for divination. In lowland Mesoamerica, the earliest numerical notation was used to indicate month and day names and to indicate periods of time (as was most often the case with surviving later Maya numerical notation). It is, of course, possible that the Shang, lowland Mesoamerican, and Bambara systems were originally used for administrative functions, but there is no material evidence to support this position. At present, then, it is impossible to identify a specific function that is correlated with the development of numerical notation.

There is another way to approach this question, which is to treat the origins of numerical notation as being the consequence of a general need for visual representational techniques, without regard to the specific functions for which these systems were used.

---

[3] The status of several of the pre-colonial West African states as 'early civilizations' is increasingly accepted by archaeologists (Trigger 1993; Connah 1987).

In four cases - the Egyptian, proto-cuneiform, Shang, and Maya systems - historical data are sufficient to conclude that numerical notation was developed just prior to, or nearly simultaneously with, the indigenous development of phonetic scripts, and this may also be true of the Harappan system. In Egypt, the numerical tags found at Abydos also provide the earliest attested instances of proto-hieroglyphs. The earliest Mesoamerican inscription (San Jose Mogoté, Monument 3) contains only the day-name "1 Earthquake". A large number of the early Shang oracle bones record numerical values (e.g. indicating sacrifices to be made). Finally, of course, the proto-cuneiform tablets that represent the earliest Mesopotamian proto-writing are no more than numerical systems combined with pictorial signs for commodities. Thus, despite my reservations about Schmandt-Besserat's (1992) arguments concerning the origins of writing (see ch. 7), I agree with her that writing often seems to emerge as an outgrowth of, or alongside, independently invented numerical notation systems. Yet it would be an error to expand this generalization into a universal law. Among the Bambara and Inka, no phonetic script was associated with the numerical notation systems that developed, and the Aztec semasiographic system was not capable of representing speech directly. While every instance of independent script development followed or accompanied the development of a corresponding numerical notation system, the converse is not true.

At present, we can say with some certainty that the independent development of numerical notation is strongly correlated with both the rise of civilizations and the independent development of scripts. Yet we do not know exactly *why* numerical notation should coincide with these developments, since it was used for different functions in different civilizations and not all civilizations developed either scripts or numerical notation systems. The pursuit of answers to this question thus requires the accumulation of new data by scholars of individual civilizations.

The Macrohistory of Numerals

The phylogenetic study of structural transformations of systems and the non-phylogenetic analysis of the replacement of systems are two powerful tools for examining diachronic patterns in numerical notation. At best, however, these tools can study relations between pairs of systems (as opposed to large regional and worldwide networks of cultural contact) and at only a single point in time (the point at which the transformation or replacement occurs). Adding in the numerous social explanations for diachronic regularities that I have just discussed does not help much. We still want to know if certain factors were more important than others at different points in time, and whether broad changes in the types of societies in the world and the nature of the interactions between them affect how numerical notation systems are invented, transformed and replaced. Thus, I will now examine the macrohistory of numerical notation by analyzing worldwide trends in the rate of invention and replacement of numerical notation systems over the past 5500 years.

As I made clear in Chapter 1, many scholars of numeration have constructed extremely simple unilinear macrohistories of numerical notation involving the gradual replacement of cumulative-additive and other 'crude' systems with ciphered-positional ones, especially the Western numerals. The diachronic regularities I have outlined do little to refute this hypothetical sequence, since ciphered and positional systems do indeed tend to replace other types. Yet the large-scale history of numeration is by no means so simple. There are macrohistorical patterns to be explained, but they are not the ones expected if the unilinear theory of the evolution of numerals were correct.

To analyse the macrohistory of numerical notation in a finely grained manner, I examined the trends in the number of systems invented and replaced at different time periods to arrive at a reasonable estimate of the number of systems in use at any given time. There are problems with this approach, which neglects the different chronologies of regions that are not in contact with one another and does not take into account the number of users of each system. The number of users of numerical notation is certainly much higher today (even when considered as a percentage of the world's population) than at any other point in history, because of high literacy rates, but, of course, those individuals are using far fewer systems than in the past. Nevertheless, the number of systems in use at any given point in time is a relatively good measure of worldwide variability among systems, and the patterns from one period to another are certainly not random fluctuations. Because it is reasonably quantifiable and completely transcends the phylogenetic level of analysis, it is useful to analyse the history of numerical notation in this additional way.

Figure 12.1 graphs the invention and extinction of all systems used over a period of 100 years or more[4], which encompasses 75 of the approximately 100 systems examined in this study, and from these figures I derive the total number of systems in use in every century from 3000 BC to 2000 AD. At first glance, the rate of invention and extinction of systems appears random throughout most of history; however, when these figures are aggregated, various patterns become clear. Five distinct phases can be identified: a) 3000 to 800 BC, when there is very little growth in the number of systems in use; b) 800 BC to BC/AD, when there is a very rapid increase in the number of systems in use; c) BC/AD to 800 AD, which is marked by relative stagnation; d) 800 to 1500, a second period of rapid increase; and e) 1500 to the present, a period of rapid and marked decline in the use of systems. Each of these phases is correlated with specific patterns of inter-cultural contact, functions for which numerical notation was or was not used, and the types of society that used numerical notation. Even though only the last period represents a relatively unified world system, I believe that these patterns accurately reflect changing socio-historical conditions influencing numerical notation. Shifting to a larger scale, it is

[4] Systems of less than 100 years duration are too short-lived to be analyzed using macrohistorical techniques and are therefore ignored. If I had included them, the only major effect on Figure 12.1 would have been that the decline after 1500 would level off in the twentieth century due to the invention in colonial contexts of many systems that were quickly abandoned or replaced.

possible to divide the history of numerical notation into two periods, the first of which is characterized by a roughly linear trend that rises steadily from 3000 BC to 1500 AD, and the second of which is a period of sharp decline from 1500 to the present.

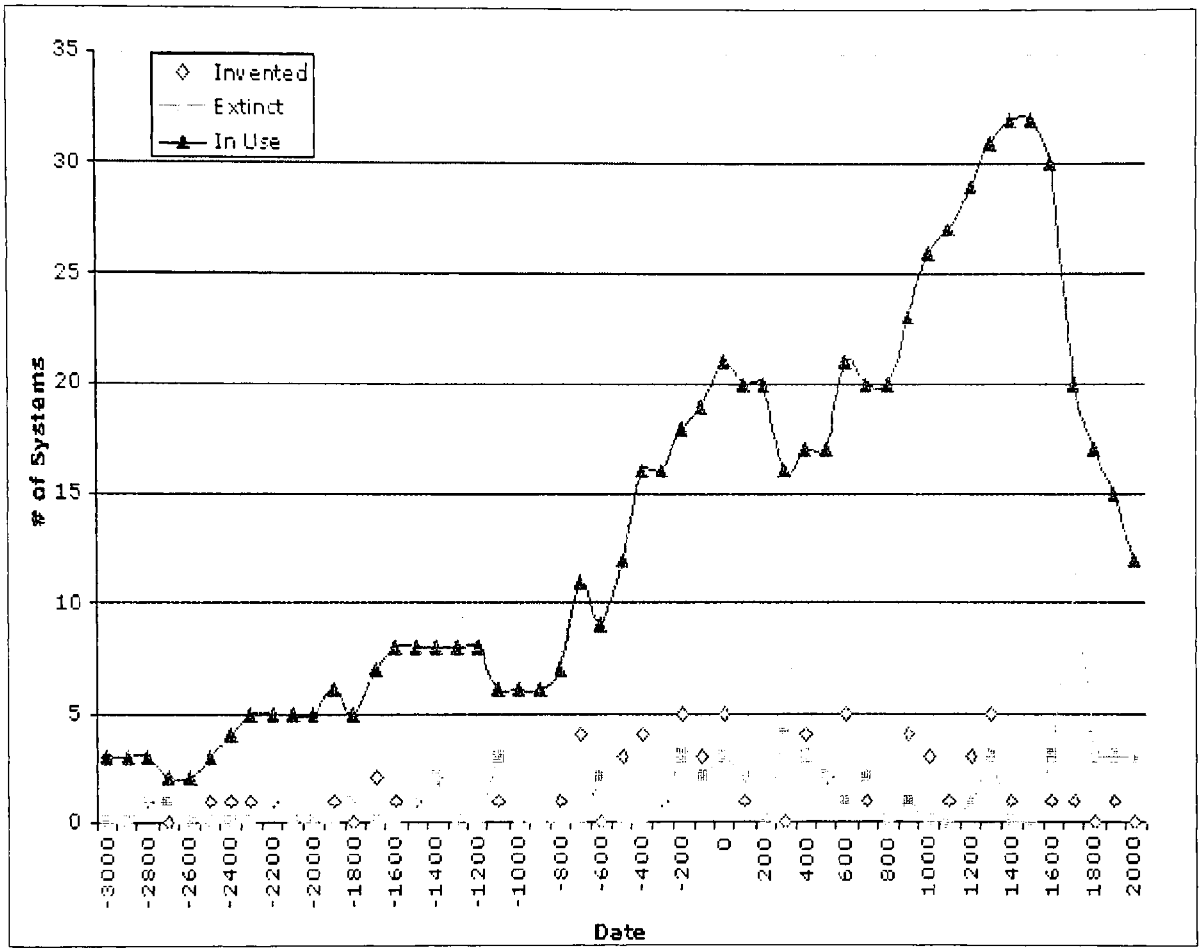


Figure 12.1: Systems in use

3000 - 800 BC: This first phase in the history of numerical notation encompasses systems invented in the civilizations of the Old World, first in Egypt and Mesopotamia, but also in the Indus Valley and China, and including systems used by secondary or peripheral civilizations (Minoan, Hittite, Eblaite, etc.). Most, but not all, the systems of this period were cumulative-additive. Numerical notation was infrequently used in inter-regional trade, as far as we can tell. The integration of such trade networks was limited; hence, the opportunities for cross-cultural contact were not as great as they would become in later periods. The contacts that did occur (between Egypt and Mesopotamia, for instance) seem not to have been conducive to the transmission of ideas about numerical notation between regions, and thus neither the invention of new systems nor the replacement of older ones was governed by such contacts. The invention of new systems during this period coincides with the development of new scripts (either endogenously or exogenously) and it is uncommon for the numerals of one script to be adopted without modification into another script. The replacement of systems during this period was largely due to the gradual transformation of older systems (proto-cuneiform → Sumerian → Assyro-Babylonian; Linear A → Linear B), and thus had no net effect on the number of systems in use. The most significant and rapid change in the use of systems during this period was in the 12th century BC, when three systems (Hittite, Ugaritic, Linear B) ceased to be used during a period of sociopolitical upheaval in the eastern Mediterranean. Overall, this was a period of slow growth and relative stability.

800 BC - BC/AD: This period might be called the 'axial age' of numerical notation systems, although it ends slightly later than Jaspers' (1953) traditional definition of that period (800 - 200 BC) as it related to the development of world religions. While I reject any teleological or mystical theories that have been associated with the 'axial age' concept, I believe that the processes involved in the rapid formation of new numerical notation systems during this period were akin to those leading to the somewhat similar but distinct world religions across Eurasia. The formation of complex networks of inter-regional trade and cultural transmission, coupled with the expansion of literate traditions into several previously non-literate or mostly non-literate regions (Italy, Greece, India, and the Levant) inspired the relatively rapid development of new scripts and corresponding numerical notation systems, including most of the systems of the Levantine and Italic families as well as many others. This period also saw the development of the first New World numerical notation systems in Mesoamerica. Cumulative-additive systems are more common than other types, but all combinations of

principle except ciphered-positional are attested. In general, political fragmentation was more typical of this period than large empires. Since each small polity or group of polities tended to develop its own script, and because of the continuation of the pattern where each new script had its own distinct numerical notation, there was a substantial increase in the rate of invention of such systems. Far fewer systems became extinct during this period than were invented, although several of the Levantine and Italic systems were relatively short-lived. Yet, at the end of this period, many of the older cumulative-additive systems of the circum-Mediterranean and Middle East were replaced by the ciphered-additive Greek alphabetic numerals, a direct consequence of the spread of Greek learning in the Hellenistic period.

BC/AD – 800 AD: In terms of the number of systems used, this period was one of stability, with episodes of invention and extinction roughly equal in frequency. New systems were invented with considerable frequency in the Alphabetic and South Asian families, derived from the Greek alphabetic and Brāhmī systems. These new systems were mostly ciphered-additive (with some multiplicative-additive systems and, towards the end of the period, ciphered-positional systems in India). In this period, ciphered systems come to outnumber cumulative systems for the first time. Many of the systems invented in this period survive to the present day (at least in limited contexts): Ethiopic, Coptic, Tamil, Sinhalese, Syriac, Arabic abjad, and, of course, the earliest Indian positional systems. In terms of replacement, it is a period when many of the cumulative systems used for millennia in Egypt and Mesopotamia were replaced with these new ciphered systems. Another important effect was the expansion of the Roman Empire, leading to the replacement of many of the cumulative systems of Europe and the Levant by Roman numerals.

800 – 1500 AD: Like the previous period, this period was one of rapid expansion in the number of systems used. Yet this was not a consequence of extraordinarily high rates of invention as much as it was of extremely low rates of replacement. New systems

continued to be invented in this period, especially in the Alphabetic and East Asian families, but also including the two ciphered-positional systems – Western and Arabic – that are most widely used today. Nevertheless, the continued use of older systems throughout this seven-century period was primarily responsible for the rise from 20 to 32 systems. This finding contradicts any simplistic notions concerning the paucity of scholarship in the Middle Ages. In this period, ciphered and multiplicative systems became much more frequent than cumulative systems in all regions except Mesoamerica, although certain Old World cumulative systems, such as the Roman numerals and Chinese rod-numerals, continued to be used quite widely.

1500 AD – present: This was the only period in history when there was a prolonged decline in the number of systems in use. This decline was particularly marked between 1550 and 1650, when no fewer than 14 systems went extinct or else were reduced to vestigial use (for instance, in archaic or strictly liturgical contexts). Particularly hard-hit were the systems of the New World, which all went extinct during this period, but many systems of the Alphabetic family were also replaced, though less dramatically, by ciphered-positional systems (Western or Arabic, depending on the region). Moreover, virtually no new systems were invented that survived for as long as 100 years. In earlier times, it was normal for each script to have its own numerical notation system (although this system was often a mere variation on its ancestor). Over the past five centuries, although local scripts have been retained and many new scripts have been invented, Western or Arabic ciphered-positional numerals have supplanted older systems and been adopted by the users of newly invented scripts, so that there is no longer anything close to a one-to-one ratio of scripts to numerical notation systems. Today, for the first time, there are more positional systems in use than additive systems – though just barely, since many ciphered-additive systems continue to be used in vestigial contexts.

The simplest explanation for this drastic decline – that it is a consequence of European imperial expansion – no doubt has some truth to it, especially for explaining the extinction of the New World numerical notation systems as a result of Spanish conquest. Similarly, the failure of the various African systems developed in the 20th century to achieve widespread acceptance is largely a product of the overwhelming social and economic dominance of users of the Western and Arabic positional numerals. Yet simply invoking imperialism as a prime mover for this decline is overly simplistic, because the rise of the imperial powers of Western Europe was a development primarily of the eighteenth and nineteenth centuries and was the consequence of a set of earlier developments. Most of the decline in the number of systems took place in the Old World between 1500 and 1650, well before the era of greatest European colonialism. Many systems that went extinct or became obsolescent in the early part of this period, such as the Glagolitic, Armenian, and Georgian numerals, did so because of the expansion of the Ottoman Empire into southeastern Europe and the Caucasus, and thus had nothing to do with European expansion (in fact, quite the opposite). For that matter, the decline of the Roman numerals and their variants (calendar numerals and Arabico-Hispanic numerals), as well as the Glagolitic and Cyrillic alphabetic systems, can hardly be explained by European conquests, since these systems were used by high-status, well-educated Europeans.

It can hardly be coincidental that the period of great decline in the number and variety of numerical notation systems in use worldwide between 1500 and 1650 corresponds to the 'long sixteenth century' demonstrated by Wallerstein (1974) to mark the formation of the capitalist world-system. The rise of capitalism and the concomitant development of superior transportation and communication technologies, as well as the conquest of the civilizations of the New World, explain the sharp falloff in the number and variability of systems. If we view numerical notation as a communication system and an administrative tool, it is not too difficult to see that a dramatic expansion in the need for communication and administration on a worldwide basis would alter dramatically the fates of the systems in use before that time.

One major reason why the reduction in numerical notation systems worldwide would corresponded to the rise of the capitalist world-system is that no earlier interregional network had nearly the same scope or strength. While there were certainly multiple 'world-systems' (in the sense of relatively closed hierarchical networks of interregional socio-economic interaction) before 1500, their ability to overwhelm older knowledge systems and cultural phenomena was not nearly as great (Abu-Lughod 1989). The late Middle Ages saw the spread of the Arabic numerals through Spain and Italy into Western Europe, but until the advent of widespread middle-class literacy in Europe, older systems such as the Roman numerals were unlikely to be replaced. Yet the capitalist world-system was an agent of an entirely different order of magnitude. By 1650, Western numerals were being introduced to new users in China, India, North and South America, and Africa, both through casual exposure in the course of economic transactions and the implementation of European secular and religious educational institutions in missions and ports of trade. The role of the Jesuits in the spread of Western numerals remains an understudied but very interesting topic. It is in no way contradictory to insist that both the spread of religious education and the rise of the capitalist world-system are involved in the replacement of older numerical notation systems. At a very basic level, since a system's perceived usefulness is related to the number and status of its users, the development of the world-system increased the number of people exposed to numerical notation and made it overwhelmingly likely that the system associated with core states would be adopted throughout the system.

Of course, the situation is somewhat more complex than a simple accounting of the number and status of the users of various systems. In analysing the role played by the rise of the world-system in the replacement of older systems by Western and Arabic positional numerals, we must examine the rise and rapid spread of the functional contexts and media in which Western numerals predominated. In the Middle Ages, both in Europe and the Islamic world, literacy and higher education were relatively restricted, and moreover in Europe literacy was strongly associated with the Latin language, which was inevitably tied to the persistence of Roman numerals. The invention of the printing press in the middle of the fifteenth century encouraged a significant increase in literacy among the middle classes of Western Europe. This technological development was concomitant with an increase in the importance and size of the middle class that in turn was associated with the development of new trade networks and shifted interest in literacy from Latin to vernacular languages. While printed books continued to use other systems for certain functions, printers were unencumbered by the tradition of Roman numeral usage of the earlier scribal tradition and frequently employed Western numerals for pagination and for representing numbers in text. The rapid spread of printed books in the sixteenth century thus ensured that readers of these works were familiarized with Western numerals, and helped overcome the stigma that had been attached to them previously as a foreign and therefore suspect innovation. The sixteenth century was the first in which printed Bibles and other religious texts in many regions began to use Western numerals alongside or in place of Roman numerals (Williams 1997). The use of dated coinage expanded dramatically starting around 1500, a function for which Roman numerals were not really suited due to the length of their numeral-phrases, and this also would have encouraged widespread familiarity with the Western numerals. The spread of coinage as a medium of international trade exposed an enormous range of individuals (many of whom were illiterate or largely so) to Western numerals. Moreover, Western numerals were more suited than either the Roman or alphabetic systems for double-entry bookkeeping, which was invented in the thirteenth century but did not become overwhelmingly popular outside Italy for a couple of centuries. Finally, the rise of modern mathematics must be considered. The need for a concise and infinitely extendable system for writing numbers helped to promote the use of ciphered-positional numerals over other principles. The problem with invoking mathematics as an explanation for the decline in systems in the modern era is that this does not explain why existing systems were replaced rather than simply altered into a ciphered-positional structure. While bookkeeping, coinage, mathematics, and printed books no doubt contributed to the ability of Western societies to dominate others, this was a later development that could not have occurred without the set of social and technological changes that had accompanied the rise of the capitalist world-system in Western Europe.

The rise of the capitalist world-system was the most significant event in the history of numerical notation. By comparison, the shift to ciphered-positional numerals and the invention of zero in medieval India, which are interpreted as being crucially important for the history of mathematics, seem relatively insignificant. Because numerical notation systems are used primarily for representation and communication, once Western European states had become core states in the world-system, it was highly desirable for the numerals associated with the administration of these states to be applied elsewhere. As more and more societies adopted Western numerals, a process of positive feedback began that accelerated their dominance, because a system with many users is more useful for communication than one with few users. Moreover, Western numerals (and other ciphered-positional systems such as the Arabic system) were very useful for a set of new and emergent functions (such as bookkeeping and mathematics) that arose in core states and aided them in maintaining their hegemonic position. Of course, other numerical notation systems continued to survive in limited contexts (primarily those not directly involved with economic concerns), while in other cases, there was considerable resistance to the new invention (particularly in China, where it has only taken hold within the last 50 years).

What can be said, then, about the prospects for the currently surviving numerical notation systems? This study is not an exercise in futurology, but I think that some provisional conclusions can be drawn from events of the past. At present, no system seems likely to replace the Western numerals as the dominant world system. Although the Arabic and various South Asian systems continue to enjoy some degree of health, as do the multiplicative-additive Chinese numerals, it seems overwhelmingly probable that they will continue to be used in their specific script traditions but not be adopted in others. Obviously, if the fortunes of the Western countries were to change dramatically vis-à-vis those of the Arabic or South Asian countries, and those countries became core states in the world-system, such a shift would be possible. Because these systems are ciphered-positional, the effort required to learn them is relatively minimal, so Western numerals might be supplanted in this way. Yet because the success of numerical notation systems is so closely intertwined with the media in which - and the functions for which - they are used, no system other than a ciphered-positional one has any real chance of worldwide acceptance for so long as computers continue to occupy their currently central role in the world's economy. As for the various alphabetic systems that have survived (Hebrew, Greek, Arabic, Cyrillic, Coptic, Syriac), because they continue to be used in extremely conservative religious texts, their complete extinction seems unlikely in the immediate future, but their expansion to new contexts seems equally improbable. Similarly, while the number of contexts in which Roman numerals are used is small, the prestige associated with them (and the practical function served by having an alternative to Western numerals) seems likely to ensure their continued use in those contexts in the foreseeable future. Yet the fortunes of all of these systems are tied to social and historical factors.

As for the invention of new systems, it is altogether premature to proclaim the end of numeration history. In the twentieth century, no fewer than six systems were invented that are not of the predominant decimal ciphered-positional structure (Bamum, Mende, Obɛri Ɔkaimɛ, Pahawh Hmong, Varang Kshiti, Iñupiaq). Even if these systems proved to be short-lived or change quickly in structure into decimal ciphered-positional systems, it seems likely that new systems will continue to be invented. One factor militating against such developments is that in the past some of these innovations were undertaken by individuals whose knowledge of Western numerals was limited (i.e. situations that might be described as stimulus diffusion). Because most people today (even illiterates) can use Western numerals, we might expect such innovations to become less frequent, or at least to be minor graphic variations on the basic decimal ciphered-positional structure. Another source of innovation in numeration might be sought in the systems designed for use in electronics or mathematics, such as binary, octal, and hexadecimal numbers, scientific (exponential) notation, or even the system of coloured bars used to designate the electrical resistivity of resistors. New systems of this sort will probably continue to be developed and employed; however, because they are useful only in limited contexts and often do not correspond with most lexical numeral systems, it is unlikely they would ever displace Western numerals. The development of systems for use in such limited contexts might expand the amount of variability among numerical notation systems worldwide, without, however, reducing the value of having a single worldwide representational system for numbers.

Finally, the prospect exists that at some point in the future, a 'post-positional' system might be developed, one that does not conform to any of the five combinations of principles that I have outlined in my typology or that violates the regularities I have described in a new and non-trivial way. One cannot predict what such a development might look like or what its cognitive advantages and disadvantages might be. What can be asserted with some certainty, however, is that unless this new system has a very significant advantage over ciphered-positional numeration for a set of specific functions, it will not be adopted on a widespread basis. I am convinced that the Western numerals are so prevalent as a representational system that they would have to be practically useless for such specific functions before any alternative system would replace them, just as the Roman numerals did not become obsolescent until the social need for a compact and extendable system arose in the sixteenth century. Even then, the demise of the

Roman numerals was hastened by the introduction of an entirely new class of users of numerical notation (the newly literate middle classes) who were not necessarily familiar with the Roman system. There is no such class of individuals today, since one can find Western numerals practically anywhere in the world. We might expect, however, that new systems - whether additive, positional, or something else - might play a role auxiliary to ciphered-positional numerals if they were perceived to be useful in particular contexts. In this way, new systems may continue to be invented and propagated, even if no system is likely to displace the Western numerals' predominance in the foreseeable future.

Chapter 13: Conclusion

> Out of the darkness, Funes' voice went on talking to me. He told me that in 1886 he had invented an original system of numbering and that in a very few days he had gone beyond the twenty-four-thousand mark. He had not written it down, since anything he thought of once would never be lost to him. His first stimulus was, I think, his discomfort at the fact that the famous thirty-three gauchos of Uruguayan history should require two signs and two words, in place of a single word and a single sign. He then applied this absurd principle to the other numbers. In place of seven thousand thirteen, he would say (for example) *Máximo Pérez*; in place of seven thousand fourteen, *The Railroad*; other numbers were *Luis Melián Lafinur, Olimar, sulphur, the reins, the whale, the gas, the caldron, Napoleon, Agustín de Vedia*. In place of five hundred, he would say *nine*. Each word had a particular sign, a kind of mark; the last in the series were very complicated ... I tried to explain to him that this rhapsody of incoherent terms was precisely the opposite of a system of numbers. I told him that saying 365 meant saying three hundreds, six tens, five ones, an analysis which is not found in the 'numbers' *The Negro Timoteo* or *meat blanket*. Funes did not understand me or refused to understand me.
>
> "Funes, the Memorious", Jorge Luis Borges (1964)

In Borges' story, the character Funes, blessed with a limitless memory, constructs an alternative system for representing numbers in which order and structure are irrelevant. In so doing, however, he creates a system whose symbols are so arbitrary as to render it useless to those of us whose memories are less prodigious than his own. What Funes has ignored - and what Borges sought to convey - is that, given human cognitive limitations, the existence of structure is necessary for the communication and retention of information in many domains of experience. Number is a phenomenon that is easily amenable to such structuring, and in fact requires it beyond a very basic level. Whether we write 7013 or 𐅆XXΔIII or Máximo Pérez is not simply a stylistic choice, but a decision that has important consequences. Structure reduces chaos to ordered simplicity and constrains a domain within well-defined and easily understandable rules. Numerical notation is useful because it imposes structure on the otherwise unstructured series of abstract natural numbers in a way that allows humans to manipulate them more effectively.

Over the past 5500 years, more than 100 different systems have been developed for representing numbers in a visual and primarily non-phonetic manner; in addition, there have been hundreds of paleographic variations of these systems. Very few systems are completely identical in structure to any other system, whether or not there are historical relations among them. There is thus considerable variability among the systems used worldwide. Even so, they are all structured by only three intraexponential and two interexponential principles, and are further constrained in the way they use bases, hybrid multiplication, phrase ordering, and arithmetical operations. Even if additional systems come to light (and it is certain that my study has not unearthed all numerical notation systems), I expect that they will fit into the typology I have constructed, because no attested system is so aberrant that it cannot be described within it. A multi-dimensional typology better reflects the various features of numerical notation systems than do one-dimensional schemes that regard the transition from additive to positional systems as the only meaningful basis for classification. This typology also lets us ask important questions about the patterns visible among attested numerical notation systems in a way that earlier typologies do not.

Even though numerical notation systems have been independently invented multiple times and have existed in a wide variety of societies across many millennia, they are easily learned and understood, and often can be interpreted even in the absence of other contextual clues. We can read Etruscan numerals even though we do not fully understand the Etruscan language. We can read Minoan numerals without being able to decipher other aspects of the Linear A script. We can read numerical values from Inka *quipus* even though our knowledge of how they were used and read is mostly lost. If there were no patterning in numerical notation – if there were no cognitive rules constraining how numbers could be written - we would not be able to perform such acts of translation. The patterns I have discussed are thus a refutation of radically relativistic notions concerning the way in which concepts are determined by culture. Recognizing that there may be significant differences in how societies think about numbers (especially in terms of number symbolism), the core of comparable features common to all numerical notation systems demonstrates that these differences are not insurmountable, and that considering numerical notation as a unitary phenomenon is warranted.

The degree of regularity exhibited by numerical notation systems shows the powerful constraints exerted by cognitive factors on certain expressions of cultural phenomena, and is thus a demonstration of psychic unity. There are numerous universals and near-universals among the world's numerical notation systems. The gap between systems that are imaginable and those that are actually attested is substantial, and can only be explained in terms of constraints imposed by human cognitive abilities. A wide range of numerical notation systems which one might conceive are never historically attested. The examination of numerical notation systems from this perspective thus allows a partial reconstruction of the mental processes of members of past societies. Given the epistemological limitations of the data from most numerical notation systems, such reconstructions are necessarily incomplete, but nonetheless important. At the same time, some of the universalistic claims that have been made regarding numerical notation are untrue, and some rules do have important exceptions. These exceptions are very important theoretically, as they allow the testing of hypotheses about the underlying causes of generalizations.

In addition to these important synchronic structural regularities, a smaller set of diachronic regularities governs patterns of invention and replacement of systems over time. These rules describe patterned connections among systems rather than the systems themselves. It is possible to determine these rules inductively because complete historical sequences of systems can be demonstrated, thus making it possible to trace phylogenetic and diffusionary relationships among systems. The universal diachronic methodology I adopted thus allows the empirical demonstration of historical relations, as opposed to other inferential techniques that require many assumptions. Yet the patterns discerned are multilinear rather than unilinear.

Because these patterns are strongly correlated with the structural principles of numerical notation systems, purely particularistic explanations for them will not suffice. It is not coincidental that cumulative and additive systems tend to be replaced over time with ciphered and positional ones. Systems can be evaluated in terms of various criteria such as conciseness, extendability, and sign-count, each of which has advantages and disadvantages. While there is no one single goal that humans universally seek to achieve when using numerical notation, a constellation of related goals can be identified, and various features of systems can be evaluated in terms of how well they reflect them. The existence of diachronic regularities and the commonalities among independent events of systemic transformation and replacement refute, or at least redefine, the commonly held anthropological dichotomy between independent invention (analogy) and diffusion (homology).

Despite the existence of synchronic and diachronic cross-cultural regularities among numerical notation systems, there is considerable evidence of the role played by social factors in determining how systems are invented, transmitted, and accepted. A decision to maximize conciseness rather than sign-count in a system, for instance, is not made on that basis alone, but in relation to one or more functions for which the system is to be used. We do not know *a priori* what specific functions will be most important, and thus we cannot evaluate how a system's users assess its utility, except through the empirical demonstration of specific contexts in which it is used. Even so, considerations other than purely structural or cognitive ones are often very important. The evaluation of a system also requires that we take into account the medium on which it is used, the linguistic affiliation of its users, the desire to emulate a powerful neighbour, and a great number of similar social factors.

Numerical notation systems are first and foremost representational systems. Their role as systems for communicating numerical information is logically prior to the specific functions for which they are used. They often arise alongside, or slightly earlier than, writing systems, and exist because their users feel a need for a visual and durable communication method. That they are primarily communication systems is extremely important for understanding their diffusion across time and space. Because they are used in inter-regional trade, the administration of colonies, missionary writings, and other contexts of intercultural communication, they can spread more rapidly than phenomena that are not communicative in function. Yet the path of transmission of numerical notation systems differs significantly from that of both writing systems and lexical numerals (the communication systems with which numerical notation can be compared most obviously). Numerical notation systems are trans-linguistic, and as such can spread in a way that is entirely divergent from patterns of script diffusion, since all scripts must to some extent represent certain phonemes and not others. Furthermore, while there are some correlations between the features of numerical notation systems and features of the lexical numerals of their inventors, a numerical notation system can be learned and used easily by speakers of other languages. The fact that number has two very different representational systems (lexical numerals and numerical notation) is very interesting from a cognitive perspective, because both are products of the interaction of cognitive processes and particular representational techniques.

While numerical notation is vitally important as a representational system, it has been largely irrelevant as a computational system, except in the recent past. A wide variety of computational techniques, including mental calculation, finger counting, tallies, and abaci, can be used for arithmetic and perform that function quite well. Numerical notation systems are used for recording the results of those computations, but are rarely used directly for calculating values. Hence, any analysis that treats computational efficiency as the prime mover behind the evolution of numerical notation is fatally flawed. The modern use of numerical notation in pen-and-paper calculation is largely an emergent function associated with the rise of mathematics and capitalism.

Moreover, with the advent of digital technologies in the past half-century, it may be that the use of numerical notation for doing arithmetic will not persist much longer.

More generally, it is possible to overemphasize the role of functionality and the importance of change in numerical notation systems. Most systems are very similar in structure to their ancestors and systems tend to be quite long-lived. Changes in systems are thus the exception rather than the rule. When such changes do occur, it is usually because of dramatic changes in the functions for which, or the social contexts in which, a system is used. Systems can be linked phylogenetically with minimal difficulty using evidence from their structural features together with evidence of cultural contact and the transmission of ideas. These phylogenies largely represent the fact that as long as a system is minimally adequate for a given set of functions, it will rarely be modified significantly, even when it is borrowed from another society. This stability is not simply a consequence of traditionalism among the users of numerical notation systems, but follows in large part from the use of numerical notation as a written communication system. In ancient societies in particular, there was little competition from other systems and little reason for a system's users to alter their behaviour.

Yet, over the past five centuries, the number of systems used worldwide has decreased dramatically and ciphered-positional systems have replaced non-ciphered and non-positional ones for most functions throughout most of the world. This is not simply a coincidence or a historically contingent event. It is the inevitable outcome of broad social changes related to the rise of capitalism in Western societies, in which the functions of numerical notation expanded to include accounting and the exact sciences. The much greater utility of the Western numerals for these functions led to their replacement of Roman numerals in core societies and the subsequent adoption of Western numerals in other societies into which Western institutions spread. Because the functions for which ciphered-positional numerical notation systems were most useful were also functions that aided Western societies to dominate others, the ciphered-positional numerals have spread essentially unopposed. Moreover, once the process of their diffusion had begun in earnest, the adoption of Western numerals in peripheral areas was a rational strategy for those who wanted to be able to communicate numerically with large numbers of powerful individuals. Yet there is no direct 'cultural selection' in favour of the structure of ciphered-positional numerals; otherwise, we might expect the identically structured Western and Tibetan numerals to have spread with equal effectiveness. The conclusion that positionality is the ultimate goal of numerical notation systems, or represents a 'perfect' development, is entirely irrelevant as an explanation of the present near-universality of ciphered-positional systems.

For many aspects of the anthropological analysis of number, data are still insufficient, and thus many interesting questions remain to be answered. In particular, our knowledge of many numerical notation systems is limited, or at best is in the hands of specialists who have not integrated their data into a synthetic framework. These data will help to fill out the phylogenies I have described and will confirm, refine, or refute the hypotheses I have offered for patterns of historical connection. We also would like to have much more data on the contexts in which systems are used, particularly relating to the connections among numerical notation, lexical numerals, and computation technologies. Just as there is no end-point in the history of numerical notation, there is no foreseeable end to the anthropological study of numerical notation. It is hoped that the present research has demonstrated the usefulness of a cross-cultural, diachronic approach to the examination of numerical notation and that these methods might be used in the analysis of other sociocultural phenomena for which historical sequences can be determined.

The history of numerical notation provides us with a 5500-year sequence of multilinear directional change in the visual representation of number. While the reconstruction of past mental processes using archaeological or historical data always runs the risk of over-interpretation, we can be confident that the observed patterns tell us something important about how people have thought about number, because numerical notation is a highly structured phenomenon. It is neither completely regular nor completely variable cross-culturally, and thus is ideal for analysing similarities and differences from both structural and social perspectives. As with any cultural feature, numerical notation systems have many historically particular idiosyncrasies that cannot be ignored, particularly when reconstructing patterns of intercultural transmission. Numerical notation is used very widely because it represents a common solution to problems of representation and communication faced in many complex societies. By understanding the functions for which systems were used, and the reasons why their users may have perceived them to be useful, we achieve a much more thorough understanding of how people think with numbers than we could from studying the systems alone.

Appendices

Appendix A: Numerical Notation Systems of the World

Table A.1: Structure of systems

Legend: Ch – Chapter; Intra – Intraexponential structure; Inter – Interexponential structure; Sub – Sub-base; Other – Additional signs; Mult – Lowest exponent at which multiplicative notation is used; Chunk – Grouping of sets of cumulative signs; SC – Sign-count; Limit – Lowest number not expressible (exponential notation used above 10 million)

System	Ch	Intra	Inter	Base	Sub	Other	Mult	Chunk	SC	Limit
Egyptian hieroglyphic	2	Cu	Ad	10		(5)		3 or 4	7	10000000
Egyptian hieratic	2	Ci	Ad	10			100000	N/A	41	1000000
Minoan (Linear A)	2	Cu	Ad	10				3	4	10000
Cretan hieroglyphic	2	Cu	Ad	10				5	4	10000
Mycenean (Linear B)	2	Cu	Ad	10				3	5	100000
Hittite hieroglyphic	2	Cu	Ad	10				5 or 9	4	10000
Urartian	2	Cu	Ad	10				5	1	10
Cypriote	2	Cu	Ad	10				3	2	100
Egyptian demotic	2	Ci	Ad	10			10000	N/A	37	100000
Meroitic cursive	2	Ci	Ad	10			1000	N/A	30	10000
Aramaic	3	Cu	Ad	10		(5), 20	100	3	7	100000
Phoenician	3	Cu	Ad	10		20	100	3	5	10000
Nabataean	3	Cu	Ad	10		4,5,20	100	3	6	1000
Kharoshthi	3	Cu	Ad	10		4,20	100	3	5	1000
Palmyrene	3	Cu	Ad	10		5,20	100	4	5	1000
Hatran	3	Cu	Ad	10		5,20	100	4	5	1000
Syriac	3	Cu	Ad	10		(2),5,20,(500)	100	4	7	1000
Etruscan	4	Cu	Ad	10	5			4	9	100000
Greek acrophonic	4	Cu	Ad	10	5			4	10	100000
South Arabian	4	Cu	Ad	10	5			4 or 9	6	10000
Lycian	4	Cu	Ad	10	5			4	5	1000
Roman classical	4	Cu	Ad	10	5	6, 500000		3 or 4	11	500000
Roman multiplicative	4	Cu	Ad	10	5		1000/100000	4	8	10^8
Arabico-Hispanic	4	Cu	Ad	10	5	40, 9		4	5	1000
Roman positional	4	Cu	Po	10	5			4	3	N/A
Calendar numerals	4	Cu	Ad	10	5			4	3	100
Berber	4	Cu	Ad	10	5			4	7	10000
Tuscan tallies	4	Cu	Ad	10	5			4	7	5000
Greek alphabetic	5	Ci	Ad	10			1000/10000	N/A	29	10000000
Hebrew alphabetic	5	Ci	Ad	10		400	1000	N/A	23	1000000
Coptic alphabetic	5	Ci	Ad	10			1000	N/A	28	1000000

Ethiopic	5	Ci	Ad	10			100	N/A	19	N/A
Gothic	5	Ci	Ad	10				N/A	27	1000
Armenian	5	Ci	Ad	10				N/A	36	10000
Georgian	5	Ci	Ad	10		10000		N/A	37	20000
Syriac (Serto)	5	Ci	Ad	10		400	1000/10000	N/A	24	10000000
Armenian (Shirakatsi)	5	Mu	Ad	10			1	N/A	12	N/A
Arabic abjad	5	Ci	Ad	10			1000	N/A	28	1000000
Glagolitic	5	Ci	Ad	10		1000,2000	1000	N/A	28	1000000
Cyrillic	5	Ci	Ad	10			1000	N/A	28	1000000
Coptic (Epakt)	5	Ci	Ad	10			1000/10000	N/A	29	10000000
Fez numerals	5	Ci	Ad	10			1000	N/A	27	1000000
Brahmi	6	Ci	Ad	10			100/1000	N/A	20	100000
Âryabhata's numerals	6	Mu	Ad	100	10	11-19, 21-25	1	N/A	42	10^{19}
Malayalam	6	Mu	Ad	10			1	N/A	12	1000000
Sinhalese	6	Ci	Ad	10			100	N/A	20	10000
Tamil	6	Mu	Ad	10			1	N/A	12	1000000
Indian	6	Ci	Po	10				N/A	10	N/A
Arabic positional	6	Ci	Po	10				N/A	10	N/A
Katapayadi	6	Ci	Po	10				N/A	33	N/A
Arabic (ghubar)	6	Ci	Po	10				N/A	10	N/A
Western	6	Ci	Po	10				N/A	10	N/A
Malayalam (modern)	6	Ci	Po	10				N/A	10	N/A
Proto-Cuneiform	7	Cu	Ad	60	10			5	6	360000
Proto-Elamite	7	Cu	Ad	10				5	5	100000
Sumerian	7	Cu	Ad	60	10			5	7	2160000
Eblaite	7	Cu	Ad	10		60	100	3	7	1000000
Assyro-Babylonian	7	Cu	Ad	10		60	100	3	6	1000000
Babylonian positional	7	Cu	Po	60	10			3	3	N/A
Mari	7	Cu	Po	10			1000	5	5	100000
Hittite cuneiform	7	Cu	Ad	10		60	100	3	7	100000
Ugaritic	7	Cu	Ad	10				3	2	100
Old Persian cuneiform	7	Cu	Ad	10				3	3	1000
Shang	8	Mu	Ad	10		20, 30, 40	1	N/A	13	100000
Rod-numerals	8	Cu	Po	10	5			4 or 5	3	N/A
Chinese classical	8	Mu	Ad	10		20, 30, 40	1	N/A	14	10^{11}
Kitan	8	Mu	Ad	10		20, 30	1	N/A	13	1000
Jurchin	8	Ci	Ad	10		11-19	100	N/A	30	100000
Chinese commercial	8	Mu	Ad	10	5	0	1	N/A	14	100000
Chinese positional	8	Ci	Po	10		10		N/A	11	N/A
Ryukyu (sho-chu-ma)	8	Cu	Ad	10	5			4	7	10000

Bar and dot	9	Cu	Ad	20	5	0		4	4	40
Maya quasi-positional	9	Cu	Po	20	5			4	3	N/A
Maya head-variants	9	Ci	Po	20	10	10		N/A	14	N/A
Aztec	9	Cu	Ad	20		100		5	4	160000
Texcocan line and dot	9	Cu	Ad	20	5			5	3	200
Texcocan (Kingsborough)	9	Mu	Ad	20	5		1	5	6	160000
Indus	10	Cu	Ad	10				4	2	100
Siyaq	10	Ci	Ad	10				N/A	45	100000
Cistercian	10	Ci	Ad	10				N/A	9	10000
Inka quipu	10	Cu	Po	10				9	1	N/A
Turkish cryptographic	10	Ci	Ad	10				N/A	12	10000
Cherokee	10	Ci	Ad	10		11-19	100	N/A	29	1000000
Bamum	10	Mu	Ad	10			1	N/A	14	100000
Bagam	10	Ci	Po	10		10		N/A	11	N/A
Mende	10	Mu	Ad	10		11-19	1	N/A	12	10000000
Bamum (mfemfe)	10	Ci	Po	10				N/A	10	N/A
Kpelle	10	Ci	Po	10		10		N/A	11	N/A
Oberi Ɔkaime	10	Ci	Po	20				N/A	20	N/A
Manding (N'Ko)	10	Ci	Po	10				N/A	10	N/A
Bete	10	Ci	Po	10		10		N/A	11	N/A
Fula (Dita)	10	Ci	Po	10				N/A	10	N/A
Pahawh Hmong (Source)	10	Mu	Ad	10		20	1	N/A	14	1000000
Fula (Adama Ba)	10	Ci	Po	10				N/A	10	N/A
Varang Kshiti	10	Ci	Ad	10				N/A	18	100
Wolof	10	Ci	Po	10				N/A	10	N/A
Pahawh Hmong (2nd Stage)	10	Ci	Po	10				N/A	10	N/A
Inupiaq	10	Cu	Po	20	5			4	3	N/A
Bambara	10	Ci	Ad	10		20		9	18	4000
Zuñi	10	Cu	Ad	10	5			4	3	100

Table A.2: History of systems

Legend: Ch – Chapter; First / Last – Earliest / latest attested use (negative numbers = BC); Dur – Duration of use

System	Ch	First	Last	Dur	Ancestor(s)	Descendant(s)	Successor(s)
Egyptian hieroglyphic	2	-3250	400	3650	Invented	Hieratic Minoan	Greek alpha Coptic
Egyptian hieratic	2	-2600	200	2800	Hieroglyphic	Demotic Meroitic	Demotic Greek alpha
Minoan (Linear A)	2	-1800	-1450	350	Egyptian hieroglyphic	Cretan Linear B Hittite	Linear B
Cretan hieroglyphic	2	-1750	-1600	150	Linear A	None	Linear A
Mycenean (Linear B)	2	-1550	-1150	400	Linear A	Etruscan?	Extinct
Hittite hieroglyphic	2	-1500	-700	800	Linear B?	Aramaic?	Assyro-Babylonian Aramaic
Urartian	2	-825	-650	175	Hittite	None	Assyro-Babylonian
Cypriote	2	-800	-200	600	Linear B Hittite Phoenician	None	Greek alpha Phoenician
Egyptian demotic	2	-750	450	1200	Hieratic	Greek alpha Brahmi	Greek alpha Coptic
Meroitic cursive	2	-250	350	600	Hieratic	None	Coptic
Aramaic	3	-750	-250	500	Eg. hieroglyphic / Bab. Common	Levantine	Greek alpha
Phoenician	3	-750	-100	650	Eg. hieroglyphic / Bab. Common	Levantine	Roman Greek alpha
Nabataean	3	-150	450	600	Aramaic	Kharoshthi	Arabic alpha
Kharoshthi	3	-100	300	400	Aramaic / Nabataean	None	Brahmi
Palmyrene	3	-50	300	350	Aramaic	Syriac	Roman Greek alpha
Hatran	3	-50	250	300	Aramaic	None	Roman Greek alpha
Syriac	3	50	600	550	Palmyrene	None	Arabic alpha
Etruscan	4	-550	0	550	Linear B? / Invented	Roman Gr. Acrophonic Tuscan	Roman
Greek acrophonic	4	-525	-50	475	Etruscan	Lycian South Arabian?	Greek alpha
South Arabian	4	-500	-100	400	Gk. Acrophonic?	Berber?	Arabic alpha
Lycian	4	-500	-300	200	Gk. Acrophonic	None	Greek alpha
Roman classical	4	-400	2000	2400	Etruscan	Runic Berber Roman variants Zuni	None
Roman multiplicative	4	-50	500	550	Roman	None	Western
Arabico-Hispanic	4	1100	1600	500	Roman	None	Western
Roman positional	4	1100	1600	500	Roman / Western	None	Western
Calendar numerals	4	1200	1600	400	Roman	None	Western

Berber	4	?	?	?	South Arabian? / Roman	None	Arabic alpha
Tuscan tallies	4	?	?	?	Etruscan?	None	Western
Greek alphabetic	5	-575	1650	2225	Demotic	Alphabetic systems Cistercian?	Arabic pos. Western
Hebrew alphabetic	5	-100	2000	2100	Greek alpha	Syrian	Western
Coptic alphabetic	5	350	1700	1350	Greek alpha	Epakt	Western
Ethiopic	5	350	2000	1650	Greek alpha	None	Western
Gothic	5	350	600	250	Greek alpha	None	Roman Greek alpha
Armenian	5	400	1650	1250	Greek alpha	Shirakatsi	Western
Georgian	5	450	1600	1150	Greek alpha	None	Western
Syriac (Serto)	5	500	2000	1500	Greek alpha / Hebrew	Arabic abjad?	Arabic positional
Armenian (Shirakatsi)	5	650	700	50	Armenian alpha	None	Armenian alpha
Arabic abjad	5	650	1150	500	Greek alpha / Syriac / Hebrew	Epakt Fez Turkish crypto.	Arabic positional
Glagolitic	5	875	1750	875	Greek alpha	None	Western
Cyrillic	5	900	1750	850	Greek alpha	None	Western
Coptic (Epakt)	5	1000	1700	700	Coptic / Arabic abjad	None	Arabic positional
Fez numerals	5	1275	1950	675	Greek alpha / Arab abjad	None	Arabic positional
Brahmi	6	-300	800	1100	Demotic? / Invented	Indian positional Malayalam Tamil Sinhalese, Âryabhata's?	Indian
Âryabhata's numerals	6	475	500	25	Brahmi?	Katapayadi	Katapayadi
Malayalam	6	500	1850	1350	Brāhmī	None	Malayalam pos.
Sinhalese	6	500	1900	1400	Brāhmī	None	None
Tamil	6	500	2000	1500	Brāhmī	None	None
Indian	6	575	2000	1425	Brāhmī	Arab positional	None
Arabic positional	6	800	2000	1200	Indian	African systems	None
Katapayadi	6	850	1550	700	Indian / Aryabhata	None	Indian
Arabic (ghubar)	6	875	2000	1125	Indian	Western	Arabic positional
Western	6	1200	2000	800	Ghubar	Cherokee Iñupiaq Chinese positional African systems	None
Malayalam (modern)	6	1850	2000	150	Malayalam/Indian	None	None
Proto-Cuneiform	7	-3200	-2900	300	Invented	Sumerian	Sumerian
Proto-Elamite	7	-3100	-2800	300	Proto-Cuneiform	None	Sumerian
Sumerian	7	-2900	-1500	1400	Proto-Cuneiform	Bab. positional Assyro-Babylonian Eblaite	Assyro- Babylonian
Eblaite	7	-2500	-2300	200	Sum. Cuneiform	Assyro-Babylonian	Assyro- Babylonian

Assyro-Babylonian	7	-2300	-200	2100	Sum. Cuneiform / Eblaite	Phoenician/Aramaic Ugaritic Hittite Old Persian	Aramaic Greek alpha
Babylonian positional	7	-2000	0	2000	Sum. Cuneiform	Greek astronomical	Greek alpha
Mari	7	-1800	-1750	50	Bab. common / positional	None	Assyro-Babylonian
Hittite cuneiform	7	-1650	-1200	450	Bab. common	None	Hittite hieroglyphic
Ugaritic	7	-1500	-1200	300	Bab. common	None	Aramaic
Old Persian cuneiform	7	-500	-350	150	Bab. common	None	Greek alpha
Shang	8	-1300	-250	1050	Invented	Chinese classical Rod-numerals?	Chinese
Rod-numerals	8	-300	1600	1900	Shang?/Invented	Chinese comm.	Chinese modern
Chinese classical	8	-250	2000	2250	Shang	Chinese comm. Chinese pos. Jurchin Kitan Ryukyu? Pahawh Hmong?	None
Kitan	8	925	1200	275	Chinese	Jurchin	Chinese Jurchin
Jurchin	8	1150	1550	400	Chinese, Kitan	None	Chinese Kitan
Chinese commercial	8	1500	2000	500	Chinese, Rod-numerals	None	None
Chinese positional	8	1600	2000	400	Chinese, Western	None	None
Ryukyu (sho-chu-ma)	8	?	?	?	Chinese?, rod-numerals?	None	Chinese
Bar and dot	9	-400	1600	2000	Invented	Maya positional Maya head variant Aztec?	Western
Maya quasi-positional	9	-50	1250	1300	Bar-and-dot	None	Western
Maya head-variants	9	300	1200	900	Maya positional	None	Western
Aztec	9	1200	1600	400	Bar-and-dot	Texcocan line+dot	Western
Texcocan line and dot	9	1500	1550	50	Aztec Pictographic	Kingsborough?	Western
Kingsborough Codex	9	1550	1550	0	Texcocan / Aztec?	None	Western
Indus	10	-2400	-1900	500	Invented	None	Extinct
Siyaq	10	900	1925	1025	Arabic lexical numerals	None	Arabic positional
Cistercian	10	1250	1650	400	Greek alpha?	None	Western
Inka quipu	10	1300	1600	300	Invented	None	Western
Turkish cryptographic	10	1800	1925	125	Arabic alpha?	None	Arabic positional
Cherokee	10	1830	1830	0	Western	None	Western
Bamum	10	1903	1921	18	Western/Arabic?	Bamum (mfemfe) Bagam?	Western/Arabic
Bagam	10	1910	1917	7	Arabic / Bamum	None	Western
Mende	10	1920	1980	60	Arabic	None	Western
Bamum (mfemfe)	10	1921	1931	10	Arabic / Bamum	None	Western
Kpelle	10	1930	2000	70	Arabic	None	None

Obɛri Ɔkaimɛ	10	1931	1970	39	Western	None	Western
Manding (N'Ko)	10	1950	2000	50	Arabic / Western	None	Western
Bete	10	1957	1957	0	Western	None	Western
Fula (Dita)	10	1958	1958	0	Arabic / Western	None	Western
Pahawh Hmong	10	1959	1970	11	Chinese?	Hmong 2nd Stage	Hmong 2nd Stage
Fula (Adama Ba)	10	1960	1960	0	Western	None	Western
Varang Kshiti	10	1960	2000	40	Indian	None	None
Wolof	10	1961	1961	0	Arabic / Western	None	Western
Hmong 2nd Stage	10	1970	2000	30	Pahawh Hmong, Western?	None	None
Inupiaq	10	1995	2000	5	Western	None	None
Bambara	10	?	?	?	Invented	None	Western
Zuñi	10	?	?	?	Invented / Roman?	None	Western

Appendix B: Synchronic and Diachronic Regularities

Axioms

A1. All numerical notation systems can represent natural numbers.

A2. All numerical notation systems have a base.

A3. All numerical notation systems use visual and primarily non-phonetic representation.

A4. All numerical notation systems are structured both intraexponentially and interexponentially.

General Regularities

G1. Any system that can represent N+1 can also represent N, where N is a natural number.

G2. All systems have a base of 10 or a multiple of 10.

G3. All systems form numeral-phrases through addition.

G4. No system forms numeral-phrases through division.

G5. All numerical notation systems are ordered and read from the highest to the lowest exponent of the base.

G6. No system uses signs for the operations used to derive the value of a numeral-phrase.

G7. The only visual features used to determine the numerical value of figures in numerical notation systems are shape, quantity, and position.

G8. There is never complete correspondence between the numeral-signs of a system and the lexical numerals of the language of the society where the system was invented.

G9. There is always some correspondence between the numeral-signs of a system and the lexical numerals of the language of the society where the system was invented.

G10. No system uses numeral-phrases that are read vertically from bottom to top.

G11. No system uses an identical representation for two different numbers.

Implicational Regularities

I1. If a system has a sub-base, the sub-base will always be a divisor of the primary base.

I2. No ciphered system has a sub-base.

I3. If a system is cumulative, it will group intraexponential signs in groups of between 3 and 5 signs.

I4. If a system is multiplicative-additive for a given exponent of its base, it will also be multiplicative-additive for all higher exponents of the base.

I5. If a system is non-multiplicative for a given exponent of its base, it will be non-multiplicative for all lower exponents of the base.

I6. Whenever the multiplicative principle is used in a system, the unit-sign or signs (multiplier) will precede the exponent-sign (multiplicand).

I7. No multiplicative system uses 1 as an exponent-sign.

I8. All multiplicative expressions involve only bases or their exponents as multiplicands.

I9. All composite multiplicands are strictly multiplicative.

Non-Universals

N1. Some additive numerical notation systems are infinitely extendable without the need to invent new signs.

N2. Some positional systems are not infinitely extendable and hence able to express any natural number.

N3. Some additive systems use a sign for zero.

N4. Some systems are not written and read in a one-directional straight line.

N5. Not all independently invented systems are cumulative-additive.

Transformational Regularities

T1. No additive system develops from a positional ancestor.

T2. No cumulative system develops from a non-cumulative ancestor.

T3. The only transformation that involves both intra- and interexponential change is the invention of multiplicative-additive systems from ciphered-positional ones, and vice versa.

T4. When one system that uses the multiplicative principle gives rise to another, the exponent above which the descendant is multiplicative is never higher than that of the antecedent.

Replacement Regularities

R1. No positional system is replaced by an additive system.

R2. No non-cumulative system is replaced by a cumulative system.

Appendix C: Social Contexts of Systemic Change

1. A system may be transformed or replaced because its structural features are disadvantageous for new social needs for which numerical notation is required.

2. A system may be adopted or rejected by individuals or groups because of the number of individuals or groups already using it.

3. A numerical notation system may be imposed on a society under conditions of political, economic, or cultural domination.

4. A numerical notation system may be invented in a region upon its being integrated into larger socio-economic networks or by elites in emulation of another society.

5. A system may be transformed or replaced if it is not compatible with the computational techniques used in a given society.

6. A system may be used for limited purposes in which it is useful to distinguish one series of numbers from another.

7. At the time of the diffusion of numerical notation into a region, the principle of the ancestral system may be adopted, but an indigenous set of numeral-signs is developed.

8. A descendant system may be structurally distinct from its ancestor because of differences in the lexical numerals associated with them.

9. In a historical context, when an established system is challenged by a new one, the older system may be defended and the interloper denigrated for cultural or political reasons.

10. A system may be borrowed or invented for use in a limited section of society to control the flow of information.

11. A system may be retained for prestige or literary purposes even after it has been supplanted by another system.

12. A system may be invented on the model of two or more existing systems.

13. A system may be transformed or replaced because of changes relating to the media on which, or the instruments with which, it is written.

14. A system used in multiple, politically independent or geographically diverse regions may diverge over time into several systems.

15. A system may diverge structurally from its ancestor due to factors related to the phonetic script of the society.

16. An existing system may be retained after its replacement for purposes for which it is more useful than the system replacing it.

17. A systemic transformation may result from factors relating to ideological subsystems of the society in which it is used.

Bibliography

Aberle, David F. 1974. Historical reconstruction and its explanatory role in comparative ethnology. In *Comparative Studies by Harold E. Driver and Essays in his Honor*, Joseph Jorgensen, ed., pp. 63-80. New Haven: HRAF Press.

Abu-Lughod, Janet. 1989. *Before European Hegemony: The World System A.D. 1250-1350.* New York: Oxford University Press.

Acharya, Subrata Kumar. 1993. The transition from the numerical to the decimal system in the inscriptions of Orissa. *Journal of the Epigraphical Society of India* 19: 52-62.

Adams, R.F.G. 1947. Obɛri Ɔkaimɛ: a new African language and script. *Africa* 17: 24-34.

Aharoni, Yohanan. 1966. The use of hieratic numerals in Hebrew ostraca and the shekel weights. *Bulletin of the American School of Oriental Research* 184: 13-19.

Allard, André. 1977. Le premier traité byzantin de calcul indien: classement des manuscrits et édition critique du texte. *Revue d'histoire des textes* 7: 57-107.

Allrik, H.L. 1954. The lists of Zerubabel (Nehemiah 7 and Ezra 2) and the Hebrew numeral notation. *Bulletin of the American Schools of Oriental Research* 136: 21-27.

Amiet, Pierre. 1966. Il y a 5000 ans les Elamites inventaient l'écriture. *Archéologia* 12: 16-23.

Anderson, W. French. 1956. Arithmetical computations in Roman numerals. *Classical Philology* 51(3): 145-150.

Anderson, W. French. 1958. Arithmetical procedure in Minoan Linear A and in Minoan-Greek Linear B. *American Journal of Archaeology* 62: 363-369.

Anderson, W. French. 1971. Arithmetic in Maya numerals. *American Antiquity* 36(1): 54-63.

Arnett, William S. 1982. *The Predynastic Origin of Egyptian Hieroglyphs.* Washington: University Press of America.

Ascher, Marcia and Robert Ascher. 1972. Numbers and relations from ancient Andean quipus. *Archive for History of Exact Sciences* 8: 288-320.

Ascher, Marcia and Robert Ascher. 1980. *Code of the Quipu: A Study in Media, Mathematics and Culture.* Ann Arbor: University of Michigan Press.

Ascher, Marcia and Robert Ascher. 1989. Are there numbers in the sky? In *Time and Calendars in the Inca Empire*, Mariusz S. Ziolkowski and Robert M. Sadowski, eds., pp. 35-48. BAR International Series 479.

Ascher, Marcia. 1991. *Ethnomathematics.* Pacific Grove, CA: Brooks/Cole.

Attié Attié, Bachir. 1975. La numération de "position" et l'arabe preislamique. *Hesperis Talmuda* 16: 7-23.

Avigad, N. 1975. A bulla of Jonathan the High Priest. *Israel Exploration Journal* 25: 8-12.

Barnett, Homer G. 1953. *Innovation: The Basis of Cultural Change.* New York: McGraw-Hill.

Barradas de Carvalho, Joaquim. 1957. Sur l'introduction et la diffusion des chiffres arabes au Portugal. *Bulletin des études portugaises* 20: 110-151.

Barthel, Thomas S. 1962. Zählweise und Zahlenglaube der Osterinsulaner. *Abhandlungen und Berichte des staatlichen Museum für Volkerkunde (Dresden)* 21: 1-22.

Barthel, Thomas S. 1971. Pre-contact writing in Oceania. In *Current Trends in Linguistics,* vol. 8, part 2, Thomas Sebeok, ed., pp. 1165-1186. The Hague: Mouton.

Basham, A.L. 1980. Early imperial India. In *The Encyclopedia of Ancient Civilizations,* Arthur Cotterell, ed., pp. 184-191. London: Penguin.

Bayley, Clive. 1882. On the genealogy of modern numerals. *Journal of the Royal Asiatic Society* 14: 335-376.

Beeston, A.F.L. 1984. *Sabaic Grammar.* Manchester: Journal of Semitic Studies.

Bendall, Cecil. 1896. On a system of letter-numerals used in South India. *Journal of the Royal Asiatic Society* 28: 789-792.

Bender, Marvin L., Sydney W. Head, and Roger Cowley. 1976. The Ethiopian writing system. In *Language in Ethiopia,* M.L. Bender, J.D. Bowen, R.L. Cooper, and C.A. Ferguson, eds., pp. 120-129. London: Oxford University Press.

Bennett, Emmett L. 1950. Fractional quantities in Minoan bookkeeping. *American Journal of Archaeology* 54: 204-222.

Bennett, Emmett L. 1996. Aegean scripts. In *The World's Writing Systems,* Peter T. Daniels and William Bright, eds., pp. 125-133. New York: Oxford University Press.

Bennett, Wendell C. 1963. Mnemonic and recording devices. In *Handbook of South American Indians,* vol. 5., Julian Steward, ed., pp. 611-619. New York: Cooper Square.

Berlin, Brent. 1968. *Tzeltal Numeral Classifiers.* The Hague: Mouton.

Bianco, Jean. 1990. Une numération de Polynésie Occidentale dans les tablettes de l'Ile de Pâques? In *Circumpacifica: Festschrift für Thomas S. Barthel,* vol. 2., Bruno Illius and Matthias Laubscher, eds., pp. 39-53. Frankfurt: Peter Lang.

Bidwell, James K. 1967 [1994]. Mayan arithmetic. In *From Five Fingers to Infinity,* Frank J. Swetz, ed., pp. 71-79.

Biella, Joan Copeland. 1982. *Dictionary of Old South Arabic, Sabaean Dialect.* Harvard Semitic Series, no. 25. Chico, CA: Scholars Press.

Birket-Smith, Kaj. 1966. The circumpacific distribution of knot records. *Folk* 8: 15-24.

Bischoff, Bernard. 1990. *Latin Palaeography.* Cambridge: Cambridge University Press.

Boas, Franz. 1940. *Race, Language and Culture.* New York: The Free Press.

Bonfante, Larissa. 1990. *Etruscan.* Berkeley: University of California Press.

Boone, Elizabeth Hill. 2000. *Stories in Red and Black: Pictorial Histories of the Aztecs and Mixtecs.* Austin: University of Texas.

Borges, Jorge Luis. 1964 [1956]. *Labyrinths.* Donald A. Yates and James E. Irby, eds. New York: New Directions Publishing Corp.

Boserup, Ester. 1965. *The Conditions of Agricultural Growth.* Chicago: Aldine.

Bowditch, Charles P. 1910. *The Numeration, Calendar Systems and Astronomical Knowledge of the Mayas.* Cambridge: Cambridge University Press.

Boyer, Carl B. 1944. Fundamental steps in the development of numeration. *Isis* 35: 153-165.

Boyer, Carl B. 1959. Note on Egyptian numeration. *Mathematics Teacher* 52: 127-129.

Boysen, S.T. and G.G. Berntson. 1989. Numerical competence in a chimpanzee *Pan troglodytes. Journal of Comparative Psychology* 103: 23-31.

Brainerd, Charles J. 1979. *The Origins of the Number Concept.* New York: Praeger.

Branigan, Keith. 1969. The earliest Minoan scripts – the pre-palatial background. *Kadmos* 8: 1-22.

Brashear, William. 1985. The myrias-symbol in CPR VII 8. *Zeitschrift für Papyrologie und Epigraphik* 60: 239-242.

Braune, Wilhelm and Ernst Ebbinghaus. 1966. *Gotische Grammatik.* Tübingen: Max Niemeyer Verlag.

Brice, William C. 1962-3. The writing system of the proto-Elamite account tablets of Susa. *Bulletin of the John Rylands Library* 45: 15-39.

Brice, William C. 1963. A comparison of the account tablets of Susa in the proto-Elamite script with those of Hagia Triada in Linear A. *Kadmos* 2: 27-38.

Brown, Donald R. 1991. *Human Universals.* Philadelphia: Temple University.

Brunschwig, Jacques and Geoffrey E.R. Lloyd, eds. 2000. *Greek Thought: A Guide to Classical Knowledge.* Cambridge, MA: Harvard University Press.

Bryce, T.R. 1976. Burial fees in the Lycian sepulchral inscriptions. *Anatolian Studies* 26: 175-190.

Bühler, Georg. 1896. *Indische Palaeographie.* Strassburg: Trubner.

Bühler, Georg. 1963 [1895]. *On the Origin of the Indian Brahma Alphabet.* Varanasi: Chowkhamba Sanskrit Series Office.

Buonamici, Giulio. 1932. *Epigrafia Etrusca.* Firenze: Rinascimento del Libro.

Burnell, A.C. 1968 [1874]. *Elements of South Indian Palaeography*. Varanasi: Indological Book House.

Burton, Michael L., Carmella C. Moore, John W.M. Whiting, and A. Kimball Romney. 1996. Regions based on social structure. *Current Anthropology* 37: 87-123.

Butterworth, Brian. 1999. *What Counts*. New York: Free Press.

Cajori, Florian. 1919. The controversy on the origin of our numerals. *Scientific Monthly* 9: 458-464.

Cajori, Florian. 1928. *A History of Mathematical Notations*. 2 vols. Lasalle, IL: Open Court.

Calame-Griaule, Genevieve. 1986. *Words and the Dogon World*. Philadelphia: Institute for the Study of Human Issues.

Cantineau, J. 1930. *Le nabatéen*. Paris: Ernest Leroux.

Cantineau, J. 1935. *Grammaire du palmyrénien épigraphique*. Le Caire: Imprimerie de l'Institut Français.

Caso, Alfonso. 1965. Mixtec writing and calendar. In *Handbook of Middle American Indians*, vol. 3, part 2, Robert Wauchope and Gordon R. Willey, eds., pp. 948-961. Austin: University of Texas.

Chabás, José and Bernard R. Goldstein. 1998. Some astronomical tables of Abraham Zacut preserved in Segovia. *Physis* 35(1): 1-11.

Chamberlain, Basil Hall. 1898. A quinary system of notation employed in Luchu on the wooden tallies termed Sho-chu-ma. *Journal of the Anthropological Institute of Great Britain and Ireland* 27: 383-395.

Cheng, David Chin-Te. 1925. The use of computing rods in China. *American Mathematical Monthly* 32: 492-499.

Childe, V. Gordon. 1956. *Society and Knowledge*. New York: Harper.

Childe, V.Gordon. 1963 [1953]. The prehistory of science: archaeological documents. In *The Evolution of Science*, Guy S. Métraux and François Crouzet, eds., pp. 34-76. New York: Mentor Books.

Clagett, Marshall. 1989. *Ancient Egyptian Science, Volume 1: Knowledge and Order*. 2 vols. Philadelphia: American Philosophical Society.

Cline, Howard F. 1966. The Oztoticpac Lands Map of Texcoco, 1540. *Quarterly Journal of the Library of Congress* 23(2): 77-115.

Closs, Michael P. 1986. The mathematical notation of the ancient Maya. In *Native American Mathematics*, Michael P. Closs, ed., pp. 291-370. Austin: University of Texas Press.

Closs, Michael P., ed. 1986. *Native American Mathematics*. Austin: University of Texas Press.

Closs, Michael. 1977. The nature of the Maya chronological count. *American Antiquity* 42: 18-27.

Closs, Michael. 1978. The Initial Series on Stela 5 at Pixoy. *American Antiquity* 43: 690-94.

Coedès, Georges. 1931. A propos de l'origine des chiffres arabes. *Bulletin of the School of Oriental Studies* 6: 323-328.

Cohen, Rudolph. 1981. Excavations at Kadesh-barnea, 1976-78. *The Biblical Archaeologist* 44: 93-107.

Colin, G.S. 1960. Abdjad. In *Encyclopedia of Islam*, vol. 1, pp. 97-98. Leiden: Brill.

Colin, G.S. 1971. Hisab al-djummal. In *Encyclopedia of Islam*, vol. 3, p. 468. Leiden: Brill.

Colin, Georges S. 1933. De l'origine grecque des "chiffres de Fès" et de nos "chiffres arabes". *Journal Asiatique* 222: 193-215.

Colville, Jeffrey. 1985. *The Structure of Mesoamerican Numeral Systems with a Comparison to Non-Mesoamerican Systems*. Ph.D., Tulane University.

Conant, Levi L. 1896. *The Number Concept*. New York: Macmillan.

Coningham, R.A.E., F.R. Allchin, C.M. Batt, and D. Lucy. 1996. Passage to India? Anuradhapura and the early use of the Brāhmī script. *Cambridge Archaeological Journal* 6(1): 73-97.

Connah, G. 1987. *African Civilizations*. Cambridge: Cambridge University Press.

Cooke, G.A. 1903. *A Text-Book of North Semitic Inscriptions*. Oxford: Clarendon.

Copans, Laurie. 1999. Israeli court rules Latin numerals can be inscribed on tombstones. Associated Press Newswire, July 7, 1999.

Cubberley, Paul. 1996. The Slavic alphabets. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 346-355. New York: Oxford University Press.

Cushing, Frank H. 1892. Manual concepts: a study of the influence of hand usage on culture-growth. *American Anthropologist* 5: 289-317.

Dalby, David. 1967. A survey of the indigenous scripts of Liberia and Sierra Leone: Vai, Mende, Loma, Kpelle and Bassa. *African Language Studies* 8: 1-51.

Dalby, David. 1968. The indigenous scripts of West Africa and Surinam: their inspiration and design. *African Language Studies* 9: 156-197.

Dalby, David. 1969. Further indigenous scripts of West Africa: Manding, Wolof and Fula alphabets and Yoruba "holy" writing. *African Language Studies* 10: 161-181.

Damerow, Peter and Robert Englund. 1989. *The Proto-Elamite Texts from Tepe Yahya* (American School of Prehistoric Research Bulletin 39). Cambridge: Harvard University Press.

Damerow, Peter. 1996. *Abstraction and Representation: Essays on the Cultural Evolution of Thinking*. Dordrecht: Kluwer.

Daniels, Peter T. and William Bright, eds. 1996. *The World's Writing Systems*. New York: Oxford University Press.

Dantzig, Tobias. 1954. *Number, the Language of Science*. New York: Macmillan.

Das, Sukumar Ranjan. 1927a. The origin and development of numerals. *Indian Historical Quarterly* 3: 97-120.

Das, Sukumar Ranjan. 1927b. The origin and development of numerals II. *Indian Historical Quarterly* 3: 356-375.

Datta, Bibhutibhusan and Avadhesh Narayan Singh. 1962 [1935]. *History of Hindu Mathematics*. Bombay: Asia Publishing House.

De Feis, L. 1898. Origine dei numeri etruschi. *Dissertazioni della Pontificia Accademia Romana di Archeologia* Series 2,7:1-19.

De Lacouperie, Terrien. 1883. The old numerals, the counting-rods and the swan-pan in China. *Numismatic Chronicle,* 3rd series, 3: 297-340.

De Odorico, Marco. 1995. *The Use of Numbers and Quantifications in the Assyrian Royal Inscriptions*. Helsinki: Neo-Assyrian Text Corpus Project.

Decourdemanche, M.J.A. 1899. Note sur quatres systèmes turcs de notation numérique secrète. *Journal Asiatique*, 9th series, 14: 258-271.

DeFrancis, John. 1989. *The Chinese Language: Fact and Fantasy*. Honolulu: University of Hawaii Press.

Dehaene, Stanislas. 1997. *The Number Sense*. New York: Oxford University Press.

Destombes, Marcel. 1987. Les chiffres coufiques des instruments astronomiques arabes. In *Marcel Destombes (1905-1983)*, Günter Schilder, Peter van der Krogt and Steven de Clercq, eds., pp. 127-140. Utrecht: HES Publishers.

Detlefsen, M., Erlandson, D., Clark, H. J., and Young, C. 1975. Computation with Roman numerals. *Archive for History of Exact Sciences* 15(2): 141-148.

Diller, Anthony. 1995. Sriwijaya and the first zeros. *Journal of the Malaysian Branch of the Royal Asiatic Society* 68(1): 53-66.

Diller, Anthony. 1996. New zeros and old Khmer. *Mon-Khmer Studies* 25: 125-132.

Diringer, David. 1949. *The Alphabet*. London: Hutchinson.

Divale, William. 1999. Climatic instability, food storage, and the development of numerical counting: a cross-cultural study. *Cross-Cultural Research* 33(4): 341-368.

Djamouri, Redouane. 1994. L'emploi des signes numériques dans les inscriptions Shang. In *Sous les nombres, le monde*, Alexei Volkov, ed., pp. 13-42. Extrême-Orient, Extrême-Occident 16. Paris: Université de Paris.

Dow, Sterling. 1952. Greek numerals. *American Journal of Archaeology* 56(1): 21-23.

Dow, Sterling. 1958. Mycenaean arithmetic and numeration. *Classical Philology* 53: 32-34.

Dreyer, Günter. 1998. *Umm el Qaab I*. Mainz: Verlag Philipp von Zabern.

Driver, Harold E. 1966. Geographical-historical versus psycho-functional explanations of kin avoidances. *Current Anthropology* 7: 131-182.

Dugast, I. and M.D.W. Jeffreys. 1950. *L'écriture des Bamum: sa naissance, son évolution, sa valeur phonétique, son utilisation*. Paris: Memoires de l'Institut Français d'Afrique Noire.

Durand, Jean-Marie. 1987. Questions de chiffres. *Mari, annales de recherches interdisciplinaires* 5: 605-610.

Duval, Rubens. 1881. *Traité de grammaire syriaque*. Paris: Vieweg.

Dwornik, Henryk. 1980-81. A 2^n-number system in the arithmetic of prehistoric societies. *Organon* 16-17: 199-222.

Ebbinghaus, Ernst. 1996. The Gothic alphabet. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 290-293. New York: Oxford University Press.

Englund, Robert K. 1996. The proto-Elamite script. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 160-164. New York: Oxford University Press.

Enwall, Joakim. 1994. *A Myth Become Reality: History and Development of the Miao Written Language*. 2 vols. Stockholm: Institute of Oriental Languages.

Evans, Gillian R. 1977. From abacus to algorism: theory and practice in medieval arithmetic. *The British Journal for the History of Science* 10(35): 114-131.

Evans, Sir Arthur. 1935. *The Palace of Minos*, vol. IV. London: Oxford University Press.

Faddegon, J.-M. 1932. Note au sujet de l'aboujad. *Journal asiatique* ser. 12, 220: 139-148.

Fairman, H.W. 1963. Two Ptolemaic numerals. *Journal of Egyptian Archaeology* 49: 179-180.

Fairservis, Walter A. 1992. *The Harappan Civilization and its Writing*. Leiden: Brill.

Falk, Harry. 1993. *Schrift im alten Indien: Ein Forschungsbericht mit Anmerkungen*. Tübingen: Narr.

Fernald, L.D. 1984. *The Hans Legacy: A Story of Science*. Hillsdale, NJ: Erlbaum.

Février, James. 1948. *Histoire de l'écriture*. Paris: Payot.

Feyel, Michel. 1932. Études d'épigraphie béotienne. *Bulletin de correspondance hellenique* 61: 217-235.

Fischer, Stephen. 1997. *Rongorongo: The Easter Island Script*. Oxford: Clarendon Press.

Flam, Jack D. 1976. Graphic symbolism in the Dogon granary: grains, time, and a notion of history. *Journal of African Studies* 3: 35-50.

Fleet, J.F. 1911a. Âryabhata's system of expressing numbers. *Journal of the Royal Asiatic Society* 1911(1): 109-126.

Fleet, J.F. 1911b. The Katapayadi system of expressing numbers. *Journal of the Royal Asiatic Society* 1911(2): 788-794.

Fossa, Lydia. 2000. Two khipu, one narrative: answering Urton's questions. *Ethnohistory* 47(2): 453-468.

Fossey, Charles, ed. 1948. *Notice sur les caracteres étrangers, anciens et modernes*. Paris: Imprimerie Nationale.

Fowler, D.H. 1999. *The Mathematics of Plato's Academy*. 2nd ed. Oxford: Clarendon.

Fowler, D.H. and E.G. Turner. 1983. Hibeh Papyrus i 27: an early example of Greek arithmetical notation. *Historia Mathematica* 10: 344-359.

Frege, Gottlob. 1953 [1884]. *The Foundations of Arithmetic*. Translation of *Die Grundlagen der Arithmetik* by J.L. Austin. Oxford: Oxford University Press.

Frei, Peter. 1976. Die Trilingue von Letoon, die lykischen Zahlzeichen und das lykische Geldsystem. *Revue suisse de numismatique* 55: 5-17.

Frei, Peter. 1977. Die Trilingue von Letoon, die lykischen Zahlzeichen und das lykische Geldsystem. *Revue suisse de numismatique* 56: 65-78.

Friberg, Jöran. 1978-9. *The third millennium roots of Babylonian mathematics*. Göteborg: University of Göteborg.

Friberg, Jöran. 1984. Numbers and measures in the earliest written records. *Scientific American* 250 (2): 110-118.

Frye, Richard N. 1973. Sasanian numbers and silver weights. *Journal of the Royal Asiatic Society* 1: 2-11.

Fukuyama, Francis. 1992. *The End of History and the Last Man*. New York: Free Press.

Fulton, Charles C. 1979. Elements of Maya arithmetic. *The Epigraphic Society Occasional Publications* 6(133): 167-180.

Fuson, Karen C. 1988. *Children's Counting and Concepts of Number*. New York: Springer-Verlag.

Gallistel, C.R. 1990. *The Organization of Learning*. Cambridge, MA: MIT Press.

Gamkrelidze, Thomas V. 1994. *Alphabetic Writing and the Old Georgian Script*. Delmar, NY: Caravan Books.

Ganay, Solange de. 1950. Graphies Bambara des nombres. *Journal de la Société des Africanistes* 20: 295-305.

Gandz, Solomon. 1931. The origin of the Ghubar numerals, or the Arabian abacus and the articuli. *Isis* 16: 393-424.

Gandz, Solomon. 1933. Hebrew numerals. *Proceedings of the American Academy of Jewish Research* 4: 53-112.

Gardiner, Sir Alan. 1927. *Egyptian Grammar*. Oxford: Clarendon.

Gardiner, Sunray Cythna. 1984. *Old Church Slavonic: An Elementary Grammar*. Cambridge: Cambridge University Press.

Gardner, Ernest A. 1888. The inscriptions. In *Naukratis, Part I*, by W.M. Flinders Petrie, pp. 54-63. London: Trübner & Co.

Gasparov, Boris. 2001. *Old Church Slavonic*. Munich: Lincom Europa.

Gasperini, Lidio. 1986. Il sistema numerale cirenaico e una nuova epigraphe dall'agorà di Cirene. *Annali della Facolta di Lettere e Filosofia, Universita di Macerata* 19: 357-366.

Geertz, Clifford. 1965. The impact of the concept of culture on the concept of man. In *New Views of the Nature of Man*, John R. Platt, ed., pp. 93-118. Chicago: University of Chicago Press.

Geertz, Clifford. 1984. Anti anti-relativism. *American Anthropologist* 86: 263-278.

Gerdes, Paulus. 1994. On mathematics in the history of sub-Saharan Africa. *Historia Mathematica* 21: 345-376.

Gibson, Craig A. and Francis Newton. 1995. Pandulf of Capua's *De calculatione:* An illustrated abacus treatise and some evidence for the Hindu-Arabic numerals in eleventh-century south Italy. *Medieval Studies* 57: 293-335.

Gibson, John C.L. 1971. *Textbook of Syrian Semitic Inscriptions*. Oxford: Clarendon.

Gillings, Richard J. 1972. *Mathematics in the Time of the Pharaohs*. Cambridge, MA: MIT Press.

Gillings, Richard J. 1978. The mathematics of ancient Egypt. In *Dictionary of Scientific Biography*, vol. 15, supplement 1, Charles Coulston Gillespie, ed., pp. 681-705. New York: Scribner.

Glautier, M.W.E. 1972. A study in the development of accounting in Roman times. *Revue internationale des droits de l'antiquité* 19: 311-343.

Goerwitz, Richard L. 1996. The Jewish scripts. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 487-497. New York: Oxford University Press.

Gokhale, Shebhana Laxman. 1966. *Indian Numerals*. Poona: Deccan College.

Goldenweiser, Alexander. 1913. The principle of limited possibilities in the development of culture. *Journal of American Folk-lore* 16: 259-290.

Gordon, Arthur E. 1983. *Illustrated Introduction to Latin Epigraphy*. Berkeley: UCLA Press.

Gordon, Cyrus. 1965. *Ugaritic Textbook*. Rome: Pontifical Biblical Institute.

Gow, James. 1883. The Greek numerical alphabet. *Journal of Philology* 12: 278-284.

Graham, J. Walter. 1969. X=10. *Phoenix* 23: 347-358.

Greenberg, Joseph H. 1978. Generalizations about numeral systems. In *Universals of Human Language*, J.H. Greenberg, ed., vol. 3, pp. 249-297. Stanford: Stanford University Press.

Greenberg, Joseph. 1975. Research on language universals. *Annual Review of Anthropology* 4: 75-94.

Greg, W.W. 1924. John of Basing's "Greek numerals". *Library (Transactions of the Bibliographic Society)*, 4th ser., 4: 53-58.

Grenfell, Bernard P. and Arthur S. Hunt, eds. 1906. *The Hibeh Papyri*, part 1. London: Egypt Exploration Fund.

Griaule, Marcel and Germaine Dieterlen. 1951. *Signes graphiques soudanais*. Paris: Hermann et Cie.

Griffith, F. Ll. 1909. *Catalogue of the Demotic Papyri in the John Rylands Library, Manchester*, vol. 3. Manchester: University of Manchester.

Griffith, F. Ll. 1916. Meroitic studies. *Journal of Egyptian Archaeology* 3: 22-30.

Griffith, F. Ll. 1925. Meroitic studies V. *Journal of Egyptian Archaeology* 11: 218-224.

Grohmann, Adolf. 1952. *From the World of Arabic Papyri*. Cairo: Al-Maaref Press.

Grube, Wilhelm. 1896. *Die Sprache und Schrift der Jučen*. Leipzig: Harrassowitz.

Guedj, Denis. 1996. *Numbers: The Universal Language*. Translated by Lory Frankel. New York: Harry N. Abrams.

Guitel, Geneviève. 1958. Comparaison entre les numérations aztèque et égyptienne. *Annales ESC* 13: 687-705.

Guitel, Geneviève. 1975. *Histoire comparée des numérations écrites*. Paris: Flammarion.

Gunn, Battiscombe. 1916. Notices of recent publications. *Journal of Egyptian Archaeology* 3: 279-286.

Gupta, R.C. 1983. Spread and triumph of Indian numerals. *Indian Journal of History of Science* 18(1): 23-38.

Haarmann, Harold. 1996. *Early Civilization and Literacy in Europe*. Berlin: Mouton de Gruyter.

Hackl, R. 1909. Merkantile Inscriften auf attischen Vasen. In *Münchener Archäologische Studien dem Andenken Adolf Furtwänglers Gewidmet*, pp. 1-106.

Haile, Getatchew. 1996. Ethiopic writing. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 569-576. New York: Oxford University Press.

Halevy, Joseph. 1875. *Etudes sabéennes*. Paris: Imprimerie Nationale.

Hallpike, C.R. 1979. *The Foundations of Primitive Thought*. Oxford: Clarendon.

Hallpike, C.R. 1986. *The Principles of Social Evolution*. Oxford: Clarendon.

Harris, Marvin. 1968. *The Rise of Anthropological Theory*. New York: Thomas Crowell.

Harris, R.A. 1905. Numerals for simplifying addition. *American Mathematical Monthly* 12: 64-67.

Harris, Roy. 1995. *Signs of Writing*. London: Routledge.

Harris, Zelig. 1936. *A Grammar of the Phoenician Language*. New Haven: American Oriental Society.

Harvey, Herbert R. 1982. Reading the numbers: variation in Nahua numerical glyphs. In *The Indians of Mexico in Pre-Columbian and Modern Times*, M.E.R.G.N. Jansen and Th. J.J. Leyenaar, eds., pp. 190-205. Leiden: Rutgers.

Harvey, Herbert R. and Barbara J. Williams. 1980. Aztec arithmetic: positional notation and area calculation. *Science* 210: 499-505.

Harvey, Herbert R. and Barbara J. Williams. 1981. L'arithmétique aztèque. *La Recherche* 126: 1068-1081.

Harvey, Herbert R. and Barbara J. Williams. 1986. Decipherment and some implications of Aztec numerical glyphs. In *Native American Mathematics*, Michael P. Closs, ed., pp. 237-260. Austin: University of Texas Press.

Hau, Kathleen. 1961. Oberi Ɔkaime script, texts and counting system. *Bulletin de l'Institut Français d'Afrique Noire* 23(1-2): 291-308.

Hau, Kathleen. 1967. The ancient writing of Southern Nigeria. *Bulletin de l'Institut Français d'Afrique Noire* 29(1-2): 150-191.

Hawkins, J. D. 2000. *Corpus of Hieroglyphic Luwian Inscriptions*. 2 vols. Berlin: de Gruyter.

Hawkins, J.D. 1986. Writing in Anatolia: imported and indigenous systems. *World Archaeology* 17(3): 363-375.

Heath, Thomas L. 1921. *A History of Greek Mathematics*. 2 vols. Cambridge: Cambridge University Press.

Hill, George Francis. 1915. *The Development of Arabic Numerals in Europe*. Clarendon Press, Oxford.

Holisky, Dee Ann. The Georgian alphabet. 1996. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 364-369. New York: Oxford University Press.

Holmes, Ruth Bradley and Betty Sharp Smith. 1977. *Beginning Cherokee*. 2nd ed. Norman: University of Oklahoma.

Hommel, Fritz. 1893. *Süd-arabische Chrestomathie*. Munich: Franz.

Hopkins, L.C. 1916. The Chinese numerals and their notational systems. *Journal of the Royal Asiatic Society* 68: 314-333, 737-771.

Houston, Stephen. 1997. A king worth a hill of beans. *Archaeology* 50(3): 40.

Høyrup, Jens. 1982. Investigations of an early Sumerian division problem, c. 2500 B.C. *Historia Mathematica* 9: 19-36.

Høyrup, Jens. 1985. Varieties of mathematical discourse in pre-modern socio-cultural contexts: Mesopotamia, Greece, and the Latin Middle Ages. *Science and Society* 49 (1): 4-41.

Hurford, James R. 1987. *Language and Number*. Oxford: Basil Blackwell.

Ifrah, Georges. 1985 [1981]. *From One to Zero: A Universal History of Numbers*. Translated by L. Bair. New York: Viking Penguin.

Ifrah, Georges. 1998. *The Universal History of Numbers*. Translated by David Bellos, E.F.Harding, Sophie Wood, and Ian Monk. New York: John Wiley and Sons.

Indraji, Pandit Bhagavanlal. 1876. On ancient Nagari numeration: from an inscription at Naneghat. *Journal of the Royal Asiatic Society, Bombay Branch* 12: 404-06.

Jaspers, Karl. 1953. *The Origin and Goal of History*. London: Routledge and Kegan Paul.

Jeffery, Lilian H. 1990. *The Local Scripts of Archaic Greece*. Oxford: Clarendon.

Jenkinson, Hilary. 1926. The use of Arabic and Roman numerals in English archives. *The Antiquaries Journal* 6: 263-275.

Johnston, Alan W. 1973. Two-and-a-half Corinthian dipinti. *Annual of the British School at Athens* 68: 181-189.

Johnston, Alan W. 1975. A South Italian numeral system. *La Parola del Passato* 30: 360-366.

Johnston, Alan W. 1979. *Trademarks on Greek Vases*. Warminster, UK: Aris and Phillips.

Johnston, Alan W. 1982. Two numerical notes. *Zeitschrift für Papyrologie und Epigraphik* 49: 205-209.

Jorgensen, Joseph G. 1979. Cross-cultural comparisons. *Annual Review of Anthropology* 8: 309-331.

Jorgensen, Joseph G., ed. 1974. *Comparative Studies by Harold E. Driver and Essays in his Honor*. New Haven: HRAF Press.

Joseph, G.G. 1991. *The Crest of the Peacock: Non-European Roots of Mathematics*. London: Tauris.

Judd, A.S. 1917. Native education in the northern provinces of Nigeria. *Journal of the African Society* 17: 1-15.

Justeson, John S. and Terrence Kaufman. 1993. A decipherment of epi-Olmec hieroglyphic writing. *Science* 259 (5102): 1703-1711.

Kak, Subhash C. 1990. The sign for zero. *Mankind Quarterly* 30: 199-204.

Kak, Subhash C. 1994. Evolution of early writing in India. *Indian Journal of History of Science* 29(3): 375-388.

Kara, György. 1996. Kitan and Jurchin. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 230-238. New York: Oxford University Press.

Kaufman, Ivan Tracy. 1967. New evidence for hieratic numerals on Hebrew weights. *Bulletin of the American Schools of Oriental Research* 188: 39-41.

Kavett, Hyman and Plyllis F. Kavett. 1975. The eye of Horus is upon you. *Mathematics Teacher* 68: 390-394.

Kaye, G.R. 1907. Notes on Indian mathematics - arithmetical notation". *Journal of the Asiatic Society of Bengal*, n.s. 3: 475-508.

Kaye, G.R. 1919. Indian mathematics. *Isis* 2: 326-356.

Kazem-zadeh, Husayn Iranshahr. 1915. *Les chiffres siyâk et la comptabilité persane*. Paris: Ernest Leroux.

Kelley, David H. 1976. *Deciphering the Maya Script*. Austin: University of Texas.

Kennedy, James G. 1981. Arithmetic with Roman numerals. *American Mathematical Monthly* 88(1): 29-32.

Keyser, Paul. 1988. The origin of the Latin numerals 1 to 1000. *American Journal of Archaeology* 92: 529-546.

King, David A. 1995. A forgotten Cistercian system of numerical notation. *Citeaux Commentarii Cistercienses* 46 (3-4): 183-217.

King, David A. 2001. *Ciphers of the Monks: A Forgotten Number-Notation of the Middle Ages*. Stuttgart: Franz Steiner Verlag.

Klein, Jeffrey J. 1974. Urartian hieroglyphic inscriptions from Altintepe. *Anatolian Studies* 24: 77-94.

Kluge, Theodor. 1937-1942. *Die Zahlenbegriffe*. 5 vols. Berlin.

Knorr, Wilbur R. 1982. Techniques of fractions in ancient Egypt and Greece. *Historia Mathematica* 9.

Koehler, O. 1951. The ability of birds to count. *Bulletin of Animal Behaviour* 9.

Krenkel, Werner. 1969. Das Rechnen mit römischen Ziffern. *Altertum* 15: 252-256.

Kroeber, Alfred L. 1948. *Anthropology*. 2nd ed. New York: Harcourt, Brace and World.

Kroeber, Alfred L. 1953. The delimitation of civilizations. *Journal of the History of Ideas* 14: 264-275.

Kroman, Erik. 1974. Taltegn. In *Kulturhistorisk Leksikon for Nordisk Middelalder*. Bd. XVIII. København: Rosenkilden og Bagger.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*. Cambridge, MA: MIT Press.

Kuttner, Robert E. 1986. Convenience does not explain the use of the Mayan duodecimal numeral system. *Perceptual and Motor Skills* 63: 930.

Kychanov, E.I. 1996. Tangut. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 228-230. New York: Oxford University Press.

Labarta, Ana and Carmen Barceló. 1988. *Numeros y cifras en los documentos Arábigohispanos*. Córdoba: Universidad de Córdoba.

Labat, René. 1952. *Manuel d'épigraphie akkadienne*. Paris: Imprimerie Nationale.

Lagercrantz, Sture. 1968. African tally-strings. *Anthropos* 63: 115-128.

Lagercrantz, Sture. 1973. Counting by means of tally sticks or cuts on the body in Africa. *Anthropos* 68: 569-588.

Lam Lay-Yong. 1986. The conceptual origins of our numeral system and the symbolic form of algebra. *Archive for History of Exact Sciences* 36: 183-195.

Lam Lay-Yong. 1987. Linkages: exploring the similarities between the Chinese rod numeral system and our numeral system. *Archive for History of Exact Sciences* 37: 365-392.

Lam Lay-Yong. 1988. A Chinese genesis: rewriting the history of our numeral system. *Archive for History of Exact Sciences* 38: 101-108.

Lambert, Joseph B., Barbara Ownbey-McLaughlin, and Charles D. McLaughlin. 1980. Maya arithmetic. *American Scientist* 68(3): 249-255.

Lang, Mabel. 1956. Numerical notation on Greek vases. *Hesperia* 25: 1-24.

Lang, Mabel. 1957. Herodotos and the abacus. *Hesperia* 26: 271-287.

Langley, James C. 1986. *Symbolic Notation of Teotihuacan*. Oxford: BAR.

Larfeld, W. 1902-07. *Handbuch der griechischen Epigraphik*. Leipzig: Reisland.

Laroche, Emmanuel. 1960. *Les hiéroglyphes hittites.* Paris: Editions du CNRS.

Laroche, Emmanuel. 1971. Les hiéroglyphes d'Altintepe. *Anatolia (Anadolu)* 15: 55-61.

Lattin, Harriet Pratt. 1933. The origin of our present system of notation according to the theories of Nicholas Bubnov. *Isis* 19: 181-194.

Lejeune, Michel. 1981. Procédures soustractives dans les numérations étrusque et latine. *Bulletin de la Societé de Linguistique de Paris* 76: 241-248.

Lemaire, André. 1977. *Inscriptions hébraiques*, vol. 1. Paris: Les Éditions du Cerf.

Lemay, Richard. 1977. The Hispanic origin of our present numeral forms. *Viator* 8: 435-462.

Levi Della Vida, G. 1934. Numerali greci in documenti arabo-spagnoli. *Rivista degli Studi Orientali* 14: 281-283.

Libbrecht, Ulrich. 1973. *Chinese Mathematics in the Thirteenth Century: the Shu-shu chiu-chang of Ch'in, Chiu-shao.* Cambridge: MIT Press.

Lidzbarski, Mark. 1898. *Handbuch der Nordsemitischen Epigraphik.* Weimar: Emil Felber.

Lieberman, Stephen J. 1980. Of clay pebbles, hollow clay balls, and writing: a Sumerian view. *American Journal of Archaeology* 84: 339-58.

Liverani, Mario. 1983. Fragments of possible counting and recording materials. In *Origini: Preistoria e Protostoria delle civiltà antiche*, part 2, Marcella Frangipane and Alba Palmieri, eds., pp. 511-521. Rome: Università degli studi la sapienza.

Loprieno, Antonio. 1986. Zahlwort. In *Lexikon der Ägyptologie*, W. Helck and W. Westendorf, eds., vol. 6, pp. 1306-1319. Wiesbaden: Harrassowitz.

Loprieno, Antonio. 1995. *Ancient Egyptian: A Linguistic Introduction.* Cambridge: Cambridge University Press.

Lounsbury, Floyd G. 1978. Maya numeration, computation, and calendrical astronomy. In *Dictionary of Scientific Biography*, Charles Coulston Gillespie, ed., vol. 15, supplement 1, pp. 759-818. New York: Scribner.

Love, Bruce. 1994. *The Paris Codex: Handbook for a Maya Priest.* Introduction by George E. Stuart. Austin: University of Texas Press.

Lowie, Robert H. 1920. *Primitive Society.* New York: Bonie and Liveright.

Lunt, Horace G. 2001. *Old Church Slavonic Grammar.* 7th rev. ed. Berlin: Mouton de Gruyter.

Mace, Ruth and Mark Pagel. 1994. The comparative method in anthropology. *Current Anthropology* 35(5): 549-564.

Macri, Martha J. 1985. The numerical head variants and the Mayan numbers. *Anthropological Linguistics* 27:46-85.

Macri, Martha J. 1996. Rongorongo of Easter Island. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 183-188. Cambridge: Cambridge University Press.

Maisels, Charles Keith. 1987. Models of social evolution: trajectories from the Neolithic to the state. *Man* (N.S.) 22: 331-359.

Mandler, G. and B.J. Shebo. 1982. Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General* 11: 1-22.

Mangalam, S.J. 1990. *Kharosthi Script*. Delhi: Eastern Book Linkers.

Marcus, Joyce. 1976. The origins of Mesoamerican writing. *Annual Review of Anthropology* 5: 35-67.

Marcus, Joyce. 1992. *Mesoamerican Writing Systems*. Princeton: Princeton University Press.

Marshack, Alexander. 1972. *The Roots of Civilization*. New York: McGraw-Hill.

Masson, Oliver. 1983. *Les inscriptions chypriotes syllabiques*. Paris: Editions E. de Boccard.

Matsuzawa, T. 1985. Use of numbers by a chimpanzee. *Nature* 315: 57-59.

Mazaheri, A. 1974. Formes 'sounnites' et formes 'chi'ites' des chiffres arabes au les avatars de chiffres indiens en islam. *Proceedings of the 13th International Congress of the History of Science* (1971, Moscow): 60-63.

McCarter, P. Kyle. 1975. *The Antiquity of the Greek Alphabet and the Early Phoenician Scripts*. Missoula, MT: Scholars Press.

McLeish, John. 1991. *The Story of Numbers*. New York: Fawcett Columbine.

Megally, Fuad. 1991. Numerical system, Coptic. In *The Coptic Encyclopedia*, Azis S. Atiya, ed., pp. 1820-1822. New York: Macmillan.

Mei Rongzhao. 1983. The decimal place-value numeration and the rod and bead arithmetics. In *Ancient China's Technology and Science*, pp. 57-65. Compiled by the Institute of the History of Natural Sciences, Chinese Academy of Sciences. Beijing: Foreign Languages Press.

Melchert, H. Craig. 1996. Anatolian hieroglyphs. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 120-124. New York: Oxford University Press.

Menninger, Karl. 1969 [1958]. *Number Words and Number Symbols*. Translation by Paul Broneer of *Zahlwort und Ziffer*. Cambridge, MA: MIT Press.

Messiha, Heshmat. 1994. Les chiffres coptes. *Le Monde Copte* 24: 25-28.

Meuret, Christophe. 1996. Le règlement de *Lamasba*: des tables de conversion appliquées à l'irrigation. *Antiquités africaines* 32: 87-112.

Mickel, Stanley L. 1981. 10,000 and more: teaching large numbers in Chinese. *Chinese Language Teachers Association Journal* 16: 83-93.

Millard, A. 1995. Strangers from Egypt and Greece - the signs for numbers in early Hebrew. In *Immigration and Emigration within the Ancient Near East,* K. van Lerberghe and A. Schoors, eds., pp. 189-194. Leuven: Peeters.

Miller, G.A. 1933. Our common numerals. *Science* 78 (2020): 236-237.

Miller, G.A. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review* 63: 81-97.

Millet, N. 1996. The Meroitic script. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 84-86. New York: Oxford University Press.

Möller, Georg. 1936 [1909-12]. *Hieratische paläographie*, 4 volumes. Osnabruck: Otto Zeller.

Mommsen, Theodor. 1965 [1909]. *Gesammelte Schriften*, vol. 7. Berlin: Weidmann.

Monod, Th. 1958. Un nouvel alphabet ouest-africain: le bété (Côte d'Ivoire). *Bulletin de l'Institut Français d'Afrique Noire*, ser. B, 20(3-4): 432-553.

Monteil, Vincent. 1951. La cryptographie chez les Maures. *Bulletin de l'Institut Français d'Afrique Noire* 13: 1257-1264.

Morley, Sylvanus. 1915. *An Introduction to the Study of the Maya Hieroglyphs*. Washington, DC: Smithsonian Institution.

Mühlisch, Alfred. 1985. Maya-Arithmetik. *Ethnographisch-archäologische Zeitschrift* 26(2): 223-239.

Mukherjee, B.N. 1993. The early use of decimal notation in Indian epigraphs. *Journal of the Epigraphical Society of India* 19: 80-83.

Mukherjee, R.N. 1977. Background to the discovery of the symbol for zero. *Indian Journal of History of Science* 12: 224-231.

Murray, Alexander. 1978. *Reason and Society in the Middle Ages*. Oxford: Clarendon.

Naroll, Raoul. 1964. On ethnic unit classification. *Current Anthropology* 5: 283-312.

Nau, F. 1910. Notes d'astronomie indienne. *Journal Asiatique* 10th series, 16: 209-225.

Naveh, Joseph. 1968. Dated coins of Alexander Janneus. *Israel Exploration Journal* 18: 20-25.

Needham, Joseph, with the assistance of Wang Ling. 1959. *Science and Civilization in China. Volume 3: Mathematics and the Sciences of the Heavens and the Earth*. Cambridge: Cambridge University Press.

Neugebauer, Otto. 1941. On a special use of the sign "zero" in cuneiform astronomical texts. *Journal of the American Oriental Society* 61: 213-215.

Neugebauer, Otto. 1957. *The Exact Sciences in Antiquity*. 2nd ed. Providence, RI: Brown University Press.

Neugebauer, Otto. 1975. *A History of Ancient Mathematical Astronomy.* 3 vols. New York: Springer-Verlag.

Nickerson, Raymond S. 1988. Counting, computing, and the representation of numbers. *Human Factors* 30 (2): 181-199.

Ninni, A.P. 1888-89. Sui segni prealfabetici usati anche ora nella numerazione scritta dai pescatori Clodiensi. *Atti del Reale Istituto Veneto di Scienze, Lettere e Arti* 7 (ser. VI): 679-686.

Nissen, Hans J., Peter Damerow, and Robert K. Englund. 1993. *Archaic Bookkeeping.* Translation of *Frühe Schrift und Techniken der Wirtschaftsverwaltung im alten Vorderen Orient* by Paul Larsen. Chicago: University of Chicago Press.

Nöldeke, Theodor. 1904. *Compendious Syriac Grammar.* Translated by James A. Crichton. London: Williams and Norgate.

Novum Testamentum in Linguam Amharicam. 1852. C.H. Blumhardt, ed. London: Watts.

O'Connor, M. 1996. The Berber scripts. In *The World's Writing Systems,* Peter T. Daniels and William Bright, eds., pp. 112-116. New York: Oxford University Press.

Oliverio, Gaspare. 1933. *Documenti antichi dell'Africa Italiana.* Bergamo.

Oppenheim, A. Leo. 1959. On an operational device in Mesopotamian bureaucracy. *Journal of Near Eastern Studies* 18: 121-128.

Palter, Robert. 1996. Black Athena, Afrocentrism, and the history of science. In *Black Athena Revisited,* Mary K. Lefkowitz and Guy Maclean Rogers, eds., pp. 209-266. Chapel Hill: University of North Carolina.

Pankhurst, Richard K.P., ed. 1985. *Letters from Ethiopian Rulers (Early and Mid-Nineteenth Century),* translated by David L. Appleyard and A.K. Irvine. Oxford: Oxford University Press.

Parpola, Asko. 1994. *Deciphering the Indus Script.* Cambridge: Cambridge University Press.

Parpola, Asko. 1996. The Indus script. In *The World's Writing Systems,* Peter T. Daniels and William Bright, eds., pp. 165-171. New York: Oxford University Press.

Paso y Troncoso, Francisco del. 1912. *Codice Kingsborough: Memorial de los Indios de Tepetlaotzoc al monarca español contra los encomenderos del pueblo.* Madrid: Fototipia de Hauser y Menet.

Payne, Stanley E. and Michael P. Closs. 1986. A survey of Aztec numbers and their uses. In *Native American Mathematics,* Michael P. Closs, ed., pp. 213-236. Austin: University of Texas Press.

Peet, T. Eric. 1923. *The Rhind Mathematical Papyrus.* London: British Museum.

Peet, T. Eric. 1931. Mathematics in ancient Egypt. *Bulletin of the John Rylands Library* 15: 409-431.

Penrose, Thomas F. 1984. *Mayan Cryptoquantum Numerations*. Franklin Park, NJ: Liberty Bell Associates.

Perny, Paul. 1873. *Grammaire de la langue chinoise orale et écrite*, vol. 1. Paris: Maisonneuve.

Perry, W.J. 1923. *The Children of the Sun*. London: Methuen.

Perry, W.J. 1924. *The Growth of Civilization*. London: Methuen.

Peruzzi, Emilio. 1980. *Myceneans in Early Latium*. Rome: Edizioni dell'Ateneo & Bizzarri.

Petrie, W.M.F. 1926. *Ancient Weights and Measures*. Warminster, UK: Aris and Phillips.

Petruso, Karl M. 1978. Marks on some Minoan balance weights and their interpretation. *Kadmos* 17: 26-42.

Petruso, Karl M. 1981. Early weights and weighing in Egypt and the Indus Valley. *Bulletin of the Museum of Fine Arts* (Boston) 79: 44-51.

Pettersson, J.S. 1999. Indus numerals on metal tools. *Indian Journal of History of Science* 34(2): 89-108.

Pettinato, Giovanni. 1981. *The Archives of Ebla: an Empire Inscribed in Clay*. Garden City: Doubleday and Company.

Piaget, Jean. 1952. *The Child's Conception of Number*. London: Routledge and Kegan Paul.

Pihan, Antoine Paulin. 1860. *Exposé des signes de numération usités chez les peuples orientaux anciens et modernes*. Paris: Imprimerie Impériale.

Pinnow, Heinz-Jürgen. 1972. Schrift und Sprache in den Werken Lako Bodras im Gebiet der Ho von Singbhum (Bihar). *Anthropos* 67: 822-857.

Pohl, Frederik. 1966. *Digits and Dastards*. New York: Ballantine.

Potts, Daniel T. 1999. *The Archaeology of Elam: Formation and Transformation of an Ancient Iranian State*. Cambridge: Cambridge University Press.

Powell, Marvin A. 1971. *Sumerian Numeration and Metrology*. Ph. D. dissertation, University of Minnesota.

Powell, Marvin A. 1972a. The origin of the sexagesimal system: the interaction of language and writing. *Visible Language* 6: 5-18.

Powell, Marvin A. 1972b. Sumerian area measures and the alleged decimal substratum. *Zeitschrift für Assyriologie* 62: 165-221.

Powell, Marvin A. 1976. The antecedents of Old Babylonian place notation and the early history of Babylonian mathematics. *Historia Mathematica* 3: 417-439.

Price, Derek J. de Solla and Leopold Pospisil. 1966. A survival of Babylonian arithmetic in New Guinea. *Indian Journal of History of Science* 1(1): 30-33.

Prinsep, James. 1838. Examination of inscriptions from Girnar in Gujerat, and Dhauli in Cuttack. *Journal of the Asiatic Society of Bengal* 7: 352-353.

Priskin, Gyula. 2002. The Eye of Horus and the synodic month. *Discussions in Egyptology* 53: 75-81.

Priskin, Gyula. 2003. Cryptic numerals on cubit rods. *Göttinger Miszellen* 192: 61-66.

Raglan, F.R.R. Somerset. 1939. *How Came Civilization?* London: Methuen.

Ratliff, Martha. 1996. The Pahawh Hmong script. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 619-624. New York: Oxford University Press.

Reineke, Walter F. 1992. Zur Entstehung der ägyptischen Bruchrechnung. *Altorientalische Forschungen* 19: 201-211.

Renou, Louis and Jean Filliozat. 1953. *L'Inde Classique: Maneul des Études Indiennes*, vol. 2. Hanoi: École Française d'Extrême-Orient.

Restivo, Sal. 1992. *Mathematics in Society and History*. Dordrecht: Kluwer.

Ritner, Robert K. 1996. Egyptian writing. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 73-84. New York: Oxford University Press.

Rogers, Everett M. 1962. *Diffusion of Innovations*. New York: The Free Press.

Rohlfs, Gerhard. 1872. Die Zahlzeichen der Rhadamser. *Ausland* 45: 695-696.

Ross, Allan S.C. 1938. *The "numeral-signs" of the Mohenjo-daro Script*. Memoirs of the Archaeological Survey of India, vol. 57. Delhi.

Rowe, John H. 1966. Diffusionism and archaeology. *American Antiquity* 31(3): 334-337.

Russell, Bertrand. 1903. *The Principles of Mathematics*. Cambridge: Cambridge University Press.

Russell, Bertrand. 1960. *Introduction to Mathematical Philosophy*. 10th ed. London: George Allen and Unwin.

Rüster, Christel and Neu, Erich. 1989. *Hethitisches Zeichenlexikon*. Wiesbaden: Harrassowitz.

Sahlins, Marshall and Elman Service, eds. 1960. *Evolution and culture*. Ann Arbor: University of Michigan Press.

Sahlins, Marshall. 1972. *Stone Age Economics*. Chicago: Aldine.

Sahlins, Marshall. 1976. *Culture and Practical Reason*. Chicago: University of Chicago Press.

Saidan, Ahmad S. 1996. Numeration and arithmetic. In *Encyclopedia of the History of Arabic Science*, vol. 2, Roshdi Rashed, ed., pp. 331-348. London: Routledge.

Salomon, Richard. 1996. Brāhmī and Kharoshthi. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 373-383. New York: Oxford University Press.

Salomon, Richard. 1998. *Indian Epigraphy*. New York: Oxford University Press.

Sandys, Sir John Edward. 1919. *Latin Epigraphy*. Cambridge: Cambridge University Press.

Sanjian, Avedis K. 1996. The Armenian alphabet. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 356-363. New York: Oxford University Press.

Sarton, George. 1936. A Hindu decimal ruler of the third millennium. *Isis* 25: 323-326.

Sarton, George. 1936. Minoan mathematics. *Isis* 24: 375-381.

Satterthwaite, Linton, Jr. 1947. *Concepts and Structures of Maya Calendrical Arithmetics*. Joint Publication No. 3, Museum of the University of Pennsylvania and the Philadelphia Anthropological Society. Philadelphia: University of Pennsylvania.

Schanzlin, G.L. 1934. The abjad notation. *The Moslem World* 24: 257-261.

Schenker, Alexander M. 1996. *The Dawn of Slavic*. New Haven: Yale University Press.

Schmandt-Besserat, Denise. 1978. The earliest precursor of writing. *Scientific American* 238(6): 38-47.

Schmandt-Besserat, Denise. 1984. Before numerals. *Visible Language* 18(1): 48-60.

Schmandt-Besserat, Denise. 1987. Oneness, twoness, threeness. *The Sciences* 27 (4): 44-49.

Schmandt-Besserat, Denise. 1992. *Before Writing*. Austin: University of Texas Press.

Schmookler, Jacob. 1966. *Invention and Economic Growth*. Cambridge, MA: Harvard University Press.

Schroder, Paul. 1869. *Die Phönizische Sprache*. Halle: Waisenhauses.

Schub, Pincus. 1932. A mathematical text by Mordecai Comtino. *Isis* 17: 54-70.

Schuhmacher, W.W. 1974. Die zoomorphen Zahlzeichen der Osterinselschrift: eine mengentheoretische Notiz. *Anthropos* 69: 271-272.

Segal, J.B. 1983. *Aramaic Texts from North Saqqara*. London: Egypt Exploration Society.

Seidenberg, A. 1986. The zero in the Mayan numerical notation. In *Native American Mathematics*, Michael P. Closs, ed., pp. 371-386. Austin: University of Texas Press.

Seidenberg, Abraham. 1960. The diffusion of counting practices. *University of California Publications in Mathematics* 3(4): 215-299.

Seidenberg, Abraham. 1986. The zero in the Mayan numerical notation. In *Native American Mathematics*, Michael P. Closs, ed., pp. 271-286. Austin: University of Texas Press.

Sen, S.N. 1971. Mathematics. In *A Concise History of Science in India*, Bose, D.M., Sen, S.N., and Subbarayappa, B.V., eds., pp. 136-212. New Delhi: Indian National Science Academy.

Sesiano, Jacques. 1989. Koptisches Zahlensystem und (griechisch-)koptische Multiplikationstafeln nach einem arabischen Bericht. *Centaurus* 32: 53-65.

Sethe, Kurt. 1916. *Von Zahlen und Zahlworten bei den alten Ägyptern, und was für andere Völker und Sprachen daraus zu lernen ist*. Strassburg: Karl J. Trübner.

Shafer, Robert. 1950. Lycian numerals. *Archiv Orientalni* 18(4): 251-61.

Shaw, Allen A. 1938-9. An overlooked numeral system of antiquity. *National Mathematics Magazine* 13: 368-372.

Shipley, Frederick W. 1902. Numeral corruptions in a ninth century manuscript of Livy. *Transactions of the American Philological Association* 33: 45-54.

Singh, A.K. 1991. *Development of Nagari Script*. Delhi: Parimal Publications.

Smalley, William A., Chia Koua Vang, and Gnia Yee Yang. 1990. *Mother of Writing: The Origin and Development of a Hmong Messianic Script*. Chicago: University of Chicago Press.

Smith, David E. and Jekuthiel Ginsburg. 1937. *Numbers and Numerals*. Washington: National Council of Teachers of Mathematics.

Smith, David E. and L. C. Karpinski. 1911. *The Hindu-Arabic Numerals*. Boston: Ginn.

Smith, Grafton Elliot. 1923. *The Ancient Egyptians and the Origin of Civilization*. London: Harper.

Soubeyran, Denis. 1984. Textes mathématiques de Mari. *Revue d'Assyriologie* 78: 19-48.

Souissi, M. 1971. Hisab al-ghubar. In *Encyclopedia of Islam*, vol. 3, pp. 468-469. Leiden: Brill.

Ste. Croix, G.E.M. de. 1956. Greek and Roman accounting. In *Studies in the History of Accounting*, A.C. Littleton and B.S. Yamey, eds., pp. 14-73. London: Sweet and Maxwell.

Steward, Julian. 1955. Cultural causality and law: a trial formulation of the development of early civilizations. In *Theory of Culture Change*, Julian Steward, ed., pp. 178-209. Urbana-Champaign: University of Illinois Press.

Stieglitz, Robert R. 1978. Minoan mathematics or music? *Bulletin of the American Society of Papyrologists* 15: 127-132.

Stone, Ruth M. 1990. Ingenious invention: the indigenous Kpelle script in the late twentieth century. *Liberian Studies Journal* 15(2): 135-144.

Stross, Brian. 1985. Maya head variant numerals: the Olmec case. *Anthropological Linguistics* 27: 1-45.

Struik, Dirk J. 1948. *A Concise History of Mathematics*. 4th ed. New York: Dover.

Struik, Dirk J. 1968. The prohibition of the use of Arabic numerals in Florence. *Archives Internationales d'Histoire des Sciences* 21: 291-294.

Struik, Dirk J. 1982. Minoan and Mycenaean numerals. *Historia Mathematica* 9: 54-58.

Subbarayappa, B.V. 1996. *Indus Script: Its Nature and Structure*. Madras: New Era Publications.

Swiggers, Pierre. 1996. Transmission of the Phoenician script to the West. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 261-270. New York: Oxford University Press.

Taisbak, C.M. 1965. Roman numerals and the abacus. *Classica et Mediaevialia* 26: 147-160.

Takashima, K. 1985. On the quantitative complement in oracle-bone inscriptions. *Journal of Chinese Linguistics* 13(1): 44-68.

Taube, Karl. 2000. *The Writing System of Ancient Teotihuacan*. Ancient America 1. Barnardsville, NC and Washington, DC: Center for Ancient American Studies.

Teeple, John E. 1931. Maya astronomy. *Contributions to American Anthropology and History* 1(2). Carnegie Institute of Washington, Publication 403.

Temple, Major R.C. 1891. Notes on the Burmese system of arithmetic. *Indian Antiquary* 20: 53-69.

Terraciano, Kevin. 2001. *The Mixtecs of Colonial Oaxaca: Ñudzahui History, Sixteenth through Eighteenth Centuries*. Stanford: Stanford University Press.

Testen, David D. 1996. Old Persian cuneiform. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 134-137. New York: Oxford University Press.

Thomas, Ivor. 1962. *Selections Illustrating the History of Greek Mathematics*, vol. 1. Cambridge, MA: Harvard University Press.

Thompson, J. Eric S. 1941. *Maya Arithmetic*. Washington, DC: Carnegie Contributions to American Anthropology and History pub. 528, vol. 403.

Thompson, J. Eric S. 1971. *Maya Hieroglyphic Writing: An Introduction*. Norman: University of Oklahoma Press.

Thomson, Robert W. 1989. *An Introduction to Classical Armenian*. Delmar, NY: Caravan Books.

Threatte, Leslie. 1980. *The Grammar of Attic Inscriptions*, volume 1 (Phonology). Berlin: de Gruyter.

Thureau-Dangin, François. 1939. Sketch of a history of the sexagesimal system. Translated by Solomon Gandz. *Osiris* 7: 95-141.

Till, Walter C. 1961. *Koptische Grammatik*. Leipzig: Veb Verlag.

Tod, Marcus Niebuhr. 1911-12. The Greek numeral notation. *Annual of the British School at Athens* 18: 98-132.

Tod, Marcus Niebuhr. 1913. The Greek numeral systems. *Journal of Hellenic Studies* 33: 27-34.

Tod, Marcus Niebuhr. 1926-7. Further notes on the Greek acrophonic numerals. *Annual of the British School at Athens* 28: 141-157.

Tod, Marcus Niebuhr. 1936-7. The Greek acrophonic numerals. *Annual of the British School at Athens* 37: 236-57.

Tod, Marcus Niebuhr. 1950. The alphabetic numeral system in Attica. *Annual of the British School at Athens* 45: 126-139.

Tod, Marcus Niebuhr. 1979. *Ancient Greek Numerical Systems*. Chicago: Ares Publishers.

Tolstoy, Paul. 1972. Diffusion: as explanation and event. In *Early Chinese Art and its Possible Influence in the Pacific Basin*, Noel Barnard, ed., vol. 3, pp. 823-841. New York: Intercultural Arts Press.

Tozzer, A.M. 1941. *Landa's Relacion de las Cosas de Yucatan*. Papers of the Peabody Museum of American Archaeology and Ethnology, Harvard University, vol. 18. Cambridge: Harvard University Press.

Trigger, Bruce G. 1989. *A History of Archaeological Thought*. Cambridge: Cambridge University Press.

Trigger, Bruce G. 1991. Constraint and freedom - a new synthesis for archeological explanation. *American Anthropologist* 93: 551-569.

Trigger, Bruce G. 2003. *Understanding Early Civilizations: A Comparative Study*. Cambridge: Cambridge University Press.

Tuchscherer, Konrad T. 1996. *The Kikakui (Mende) Syllabary and Number Writing System*. Ph.D. dissertation, London School of Oriental and African Studies.

Tuchscherer, Konrad T. 1999. The lost script of the Bagam. *African Affairs* 98: 55-77.

Turner, Eric G. 1975. Four obols a day men at Saqqara. In *Le monde grec: Hommages à Claire Préaux*, J. Bingen, G. Cambier, and G. Nachtergael, eds., pp. 573-577. Bruxelles: Éditions de l'Université de Bruxelles.

Tylor, E.B. 1958 [1871]. *Primitive Culture*. 2 vols. New York: Harper and Brothers.

Urcid Serrano, Javier. 2001. *Zapotec Hieroglyphic Writing*. Studies in Pre-Columbian Art and Archaeology, no. 34. Washington, DC: Dumbarton Oaks Research Library.

Urton, Gary. 1997. *The Social Life of Numbers*. Austin: University of Texas Press.

Urton, Gary. 1998. From knots to narratives: reconstructing the art of historical record keeping in the Andes from Spanish transcriptions of Inka *khipus*. *Ethnohistory* 45(3): 409-437.

Urton, Gary. 2001. A calendrical and demographic tomb text from northern Peru. *Latin American Antiquity* 12(2): 127-147.

Vaillant, André. 1948. *Manuel du vieux slave*. Paris: Institut d'Études Slaves.

Vaillant, G.C. 1950. *The Aztecs of Mexico*. Harmondsworth: Pelican Books.

van den Brom, Lourens. 1969. Woher stammt das 60-system? *Janus* 56: 210-214.

Van der Waerden, B.F. 1963. *Science Awakening*. Translated by Arnold Dresden. New York: John Wiley.

Van der Waerden, B.L. 1983. *Geometry and Algebra in Ancient Civilizations*. Berlin: Springer-Verlag.

Ventris, Michael and John Chadwick. 1973. *Documents in Mycenean Greek*. 2nd ed. Cambridge: Cambridge University Press.

Verma, Thakur Prasad. 1971. *The Palaeography of Brahmi Script in North India*. Varanasi: Siddharth Prakashan.

Volkov, Alexei. 1994. Large numbers and counting rods. In *Sous les nombres, le monde*, Alexei Volkov, ed., pp. 71-91. Extrême-Orient, Extrême-Occident 16. Paris: Université de Paris.

Vycichl, Werner. 1952. Das berberische Ziffernsystem von Ghadames und sein Ursprung. *Rivista degli Studi Orientali* 27: 81-83.

Wallerstein, Immanuel. 1974. *The Modern World-System, Vol. 1: Capitalist Agriculture and the Origins of European World-Economy in the Sixteenth Century*. New York: Academic Press.

Wallis, Faith, ed. 1999. *The Reckoning of Time*, by the Venerable Bede. Liverpool: Liverpool University Press.

Wallis, Wilson D. 1925. Diffusion as a criterion of age. *American Anthropologist* 27: 91-99.

Was, Daniel A. 1971. Numerical fractions in Minoan Linear script A. *Kadmos* 10: 35-51.

Wassén, Henry. 1931. The ancient Peruvian abacus. In *Comparative Ethnographical Studies*, Erland Nordenskiöld, ed., vol. 9, pp. 189-205. New York: AMS Press.

Wells, David. 1986. *The Penguin Dictionary of Curious and Interesting Numbers*. London: Penguin.

White, Leslie. 1949. *The Science of Culture*. New York: Farrar, Strauss.

White, Leslie. 1959. *The Evolution of Culture*. New York: McGraw-Hill.

Whiting, Robert M. 1984. More evidence for sexagesimal calculations in the third millennium B.C. *Zeitschrift für Assyriologie* 74: 59-66.

Williams, Barbara J. and H. R. Harvey. 1997. *The Códice de Santa María Asunción*. Salt Lake City: University of Utah Press.

Williams, Jack. 1997. Numerals and numbering in early printed English Bibles and associated literature. *Journal of the Printing Historical Society* 26: 5-13.

Wilson, Nigel. 1981. Miscellanea paleographica. *Greek, Roman and Byzantine Studies* 22: 395-404.

Woepcke, F. 1863. Mémoire sur la propagation des chiffres indiens. *Journal Asiatique*, 6th ser, 1: 27-79, 234-290, 442-459.

Woodruff, Charles E. 1994 [1909]. The evolution of modern numerals from ancient tally marks. In *From Five Fingers to Infinity*, Frank Swetz, ed. Reprinted from *American Mathematical Monthly* 16: 125-133. Chicago: Open Court.

Worm, Ole. 1643. *Fasti Danici*. Copenhagen: Hafnia.

Wroth, Warwick. 1966. *Imperial Byzantine Coins*. Chicago: Argonaut.

Wynn, Karen. 1992. Addition and subtraction by human infants. *Nature* 358: 749-751.

Yadin, Yigael. 1961. Ancient Judaean weights and the date of the Samaria ostraca. *Scripta Hierosolymita* 8: 9-25.

Zabilka, Ivan Lee. 1968. *Ancient Near Eastern Number Systems and their Relationship to the Hebrew Number System during the Biblical Period*. M.Th. thesis, Asbury Theological Seminary.

Zaslavsky, Claudia. 1973. *Africa Counts: Number and Pattern in African Culture*. Boston: Prindle, Webber & Schmidt.

Zhang, Jiajie and Donald A. Norman. 1995. A representational analysis of numeration systems. *Cognition* 57: 271-295.

Zide, Norman. 1996. Scripts for Munda languages. In *The World's Writing Systems*, Peter T. Daniels and William Bright, eds., pp. 612-618. New York: Oxford University Press.

Zimansky, Paul. 1993. Review of Denise Schmandt-Besserat, *Before Writing*. *Journal of Field Archaeology* 20: 513-517.