Chapter 6
Reflecting on the What and Why of Whole Number Arithmetic: A Commentary on Chapter 5

Roger Howe

6.1 Introduction

Whole number arithmetic is a basic part of mathematics education everywhere, and there is a tendency, among mathematicians especially, but also I believe among mathematics educators, to think of it as 'easy' or 'simple', and in comparison to later parts of the curriculum – fractions, algebra, etc. – this attitude has some validity.

But, even WNA is not simple! At least, if history can be a guide to what is simple and what is hard, then Chap. 5 shows that our current, essentially universal, formulation of arithmetic in terms of base-ten place value notation (PVN) is not simple. It is the product of a very long historical development, some parts of which are still little understood. PVN has won its worldwide acceptance and use, not because it is simple, but for the power it brings to numeration and calculation; because once you understand it – or at least, can work with it – many other things are simple. Although PVN has that characteristic feature of a great idea – that after you learn it, you can't imagine not knowing it – it presents considerable obstacles to the learner, and many students master it only partially. PVN is not obvious. Chapter 5 gives us some perspective on why this is so.

R. Howe (✉)
Yale University, New Haven, CT, USA
e-mail: roger.howe@yale.edu

M.G. Bartolini Bussi, X.H. Sun (eds.), *Building the Foundation: Whole Numbers in the Primary Grades*, New ICMI Study Series,
https://doi.org/10.1007/978-3-319-63555-2_6

6.2 Problematics of Place Value Notation

As we learn in Chap. 5, the ancient or classical civilisations – Mesopotamian (Sumerian, Babylonian, etc.), Egyptian, Chinese, Indus Valley, Greek and Roman – although they used a lot of mathematics and did complicated calculations, none of them developed positional notation (although, in some sense, the Chinese came close). It seems to have come into being in South or Southeast Asia, perhaps not until the early centuries of the CE, as a result of interactions between Chinese merchants and various neighbours of China, more specifically as a result of trying to adapt Chinese computational methods to written form. The Chinese certainly had a highly developed base-ten arithmetic, and they already had many of the components of place value notation, especially as embedded in their computational tools, but their written numbers included explicit symbols for each base-ten unit. They had no symbol for zero; if the multiple of a particular unit needed to express a given number was zero, the unit was just omitted from the writing. The Romans used a base-ten system, but their notation was more or less what is designated in Sect. 5.2.3.2 as a system of 'additive type', with no digits. They had separate symbols for each base-ten unit and also one for five of each unit, and they used subtraction as well as addition to express numbers. Although we think of the classical Greeks as strong mathematicians and they did marvellous things in geometry, their conception of number was not as advanced. The Greek numeration system was quite ad hoc and limited. It used combinations of up to three letters to express numbers up to 999. (As related in a draft version of Chap. 5, the Hebrew and early Arabic systems were similar – was this perhaps a common legacy from Phoenician practice, as with their alphabets themselves?) So the late development of place value notation, along with the gradual evolution of the computational algorithms after it was invented, provides convincing historical evidence that PVN is far from obvious.

Carl Friedrich Gauss (1777–1855) is often named as the best post-Renaissance European mathematician. Although pure mathematicians today revere him especially for his contributions to number theory, he also was an accomplished applied mathematician. He did massive calculations by hand, including a determination of the orbital motion of the large asteroid/dwarf planet *Ceres*, from only a very few observations of its position. His prediction of where in the sky to find it again led to its rediscovery and its establishment as a notable member of the solar system.

Gauss's experiences with computation made him highly sensitive to the virtues of place value notation. He is quoted (Eves 1969) (Newman 1956, p. 328) as saying:

> The greatest calamity in the history of science
> was the failure of Archimedes
> to invent positional notation.

Here, the main point is that even Archimedes, perhaps the best of classical Greek mathematicians, had not come up with the idea of place value notation. This was not because Archimedes never thought about large numbers: he wrote a paper called *The Sand Reckoner*, whose theme was precisely how to express large numbers (such as the number of grains of sand in all the beaches of the world), and which discussed ways of constructing and naming such numbers. Nevertheless, the concept of an efficient, all-purpose, unlimited system for writing numbers, such as is provided by place value notation, did not figure in his proposals. One might wonder whether the very limited ad hoc system (described above) in common use by classical Greeks and, in particular, its lack of use of multiplication, somehow inhibited Archimedes' thinking.

6.3 Algebraic Structure and the Power of Place Value Notation

If PVN was so highly valued even by Gauss, it is worth looking at carefully. As noted in Sect. 5.2.1, positional notation harnesses many concepts, including essentially all of basic polynomial algebra, in the service of simply writing numbers. Moreover, it owes it computational efficacy to its compatibility with algebraic structure. It also relies on numerous conventions that must be understood by the reader in order to interpret base-ten numbers correctly. Three different interpretations or elaborations of place value notation are mentioned in Sect. 5.2.3.6 of Chap. 5 as having been taught in France in recent years (in addition to the 'academic theory' promoted by the New Math). My paper for the Study Conference (Howe 2015) describes a slightly more elaborated sequence of interpretations, five in all, that includes all of these in a single developmental scheme that might be taken as five stages of place value understanding and that could describe the progression of a learner through a carefully designed arithmetic curriculum. They are illustrated by the following sequence of equations:

$$\begin{aligned} 456 &= 400 + 50 + 6 \\ &= 4 \times 100 + 5 \times 10 + 6 \times 1 \\ &= 4 \times (10 \times 10) + 5 \times 10 + 6 \times 1 \\ &= 4 \times 10^2 + 5 \times 10^1 + 6 \times 10^0 \end{aligned}$$

The first interpretation of '456' mentioned in Sect. 5.2.3.6 is more or less the way '456' is read out loud in English. (Apparently, the practice is different in France, with the base-ten units suppressed.) The second and third interpretations agree with the second and third stages above.

The second stage breaks the number up into a sum of pieces. In terms of PVN, each of these pieces involves only one non-zero digit. This stage displays the basic strategy of expansion in a given base: every number can be expressed as a *sum* of pieces of a special kind. Remarkably, there appears to be no standard short name for these pieces in the mathematics or mathematics education literature. For present purposes, we will refer to them as *base-ten pieces*.

Thus, the base-ten pieces of 456 are 400, 50 and 6. This explicit decomposition of a number into the sum of its base-ten pieces is often presented in US classrooms under the name *expanded form*.

The base-ten pieces of a number themselves have substantial structure, which are to be understood in terms of multiplication. The third to fifth stages of PVN reveal successive features of this multiplicative structure.

The first aspect of this structure is that each base-ten piece is a multiple of an even more special quantity, a *base-ten unit*. These are the base-ten pieces whose non-zero digit is just 1. A general base-ten piece is a multiple of a base-ten unit, and the non-zero digit tells us what that multiple is. Thus, 400 = 4 × 100, 50 = 5 × 10 and 6 = 6 × 1. Thus, the base-ten units here are 100, 10 and 1.

With this terminology, we can say:

Each base-ten piece is a digit times a base-ten unit.

This is what the third stage is telling us. It might be called the 'second expanded form'.

We should not stop with this stage. The base-ten units themselves have multiplicative structure, and this structure is really the key to the efficacy of the idea of base-ten expansion, so it should be made explicit.

Each base-ten unit is itself a product: a repeated product of 10s. The base-ten unit 1 indicates the basic unit of the whole system. It stands for whatever quantity you are counting. As revealed already by prehistoric tally systems, all whole numbers are obtained by iterating the basic unit sufficiently many times. The next base-ten unit, 10, is the key to the whole system. It reveals the *base* or grouping ratio: each successive base-ten unit is obtained by combining 10 of the previous unit. Thus, 10 = 10 × 1 is the first base-ten unit beyond unity. The next base-ten unit is 10 × 10 = 100. The next one is 10 × 100 = 1000, and on and on, for as long as we need to go. For most everyday purposes, we don't have to go too far: since each base-ten unit is ten times the next smaller one, these numbers get large fast! In our example, we only need the first three units – 1, 10 and 100. The standard Greek, Hebrew and Arabic notational system was content with representing numbers only up to 999. Roman numerals went somewhat farther but not much.

Finally, the last expression summarises the previous one by expressing the iterated products of 10 with itself in exponential notation. It is the 'academic theory' (see Sect. 5.2.3.6) specialised to base ten. The expression bears strong resemblance to the way polynomials are written in algebra and indeed can be thought of as expressing the given number as a 'polynomial in 10', with the understanding that the 'coefficients' of the 'polynomial', i.e. the digits in the first expression, are all whole numbers less than 10.

From an educational perspective, a key point to realise about the five stages of place value is that although to a mature understanding all these expressions are more or less obviously equivalent, each stage represents a substantial intellectual advance on the previous one. For example, exponential notation is justified using the associative rule for multiplication, which is arguably the deepest of the rules of arithmetic. Correspondingly, exponential notation is usually not introduced until late elementary school – not in the primary grades – and well after PVN has already been in use by students with 3–5-digit (or more) numbers. (However, this late introduction is probably not done to allow a principled discussion of the role of the associative rule in defining exponential notation!) Taken all together, understanding the five stages implies a lengthy intellectual development that requires, if it is achieved at all, the full span of elementary education.

6.4 Possible Lessons for Education

All this is well known to mathematics educators, but it seems worthwhile to rehearse it again here, for several reasons.

First, it can help us to see to what may be some omissions in educational discourse and in WNA instruction. The first example of this would be the lack of a short name for talking about the basic building blocks of the notation, what we are here calling the 'base-ten pieces'. A short name would facilitate discussion of them and their role in PVN, thereby promoting conceptual understanding of PVN.

Further on, we should ask to what extent the full structure revealed in the fifth expression, that is, the algebraic structure implicit in PVN, is made clear to students. Some evidence suggests that, in the United States, even the third stage, the 'second expanded form', does not become part of the thinking of a large segment, perhaps a majority, of students (Thanheiser 2009, 2010). We should consider how to structure our curricula so that the ideas embodied in the five stages of place value, and the algebraic structure underlying it, are absorbed by students. This would be consistent with the 'higher-order thinking' mantra of twenty-first century education.

Second, the unlimited nature of place value notation is brought home when we use exponential notation to express base-ten pieces as $d \times 10^k$, where we understand that k can be any whole number. The WNA curriculum as we have inherited it from previous centuries might be described as 'small-number-centric' (which was appropriate for many of its primary justifications, e.g. 'shopkeeper arithmetic'). It starts with single digits, proceeds to two-digit numbers, then to three- and four-digit numbers, and then perhaps with a little attention to five- and six-digit numbers, tends to think its job is done. This is probably fine if the main goal of instruction is to enable people to calculate correctly with medium-sized numbers, but it does not convey a sense of the system as a whole. This small-number focus may be part of the reason many people have little appreciation for the difference between a million and a billion, thinking of both of them simply as 'very big numbers'.

But one can argue that today, it is an important civic skill to understand the difference between a million and a billion. To understand, for example, that a billionaire is equivalent to a thousand millionaires.[1] If a billionaire spends 10 million dollars on a house and takes a million dollar vacation each year, he still has about $950 million left to do other things with. When Bill Gates built his house, everyone

[1] This is in the USA. In England, 'billion' means 'million million'.

was agog that he spent 40 million dollars on it. But Gates' net worth at the time was 40 billion dollars, so he was spending 0.1% of his net worth on the house. What kind of a house could you buy for 0.1% of your net worth? In this context, it should also be taken into account that rich people do not keep their money under their pillows, they invest it: their money makes more money. If Gates' fortune was increasing at just 1% per year (in fact, it was increasing much faster), he was richer after he paid for the house than when he signed the contract to have it built.

Furthermore, to discuss intelligently economic constructs such as gross national product, billions are not enough; one needs at least trillions. For example, the GDP of the United States in 2015 was about 18 trillion dollars.

Discussion of issues like climate change involves comparing numbers that are substantially larger than this. For example, how big is Earth's atmosphere? The weight of the atmosphere is about 15 pounds per square inch (atmospheric pressure) times the area of Earth in square inches. How many pounds is that? Computations like this one also bring home the point that, not only would it be very cumbersome to attempt to compute all the digits in a large number, it would be a waste of time.

With some effort, one can compute by hand that the number of square inches in a square mile is 4,014,489,600. (This computation does not exceed the capacity of many hand calculators, so a modern child can find it with a few keystrokes – if she/he knows what to do.) Then all we have to do is multiply this by the area of Earth in square miles. But how accurately can we know this? Do we want to try to find the exact true area of Earth in all its glorious roughness? This does not even make sense. Most of the surface of Earth is water, which is constantly being jostled by the wind to form waves. Waves change the surface area of water, sometimes quite drastically. (Think of 'The Great Wave off Kanagawa' in Hokusai's woodblock print. Some would put this forward as an example of a fractal, with infinite surface area.) A simpler approach might be to pretend that Earth is a sphere and use the formula $A = 4\pi r^2$ for its area. To carry out this strategy, you have to confront the facts that (i) Earth is not in fact a sphere and, in particular, (ii) its 'radius' is not exactly defined. In fact, the 'radius of Earth' does not make sense to much more accuracy than ±5 miles.[2] Since it is approximately 4000 miles, this means that we know the radius of Earth to less than three significant figures. Keeping in mind the principle that a product is only known as accurately as the least accurate of its factors, it does not make sense to report more than the three largest base-ten pieces in describing the area of Earth or the weight of the atmosphere. So our lovely calculation above of the number of square inches in a square mile could (and should) just be replaced by 'approximately 4 billion'. The corresponding figure for the area of Earth in square inches is quite adequately represented by 800,000,000,000,000,000 or 800 quadrillion (in US numeration).

To successfully teach children to comprehend and work with numbers this large, we have to get away from focusing on the digits and work hard to understand the sizes of the pieces: to focus more attention on the base-ten pieces, especially the

[2]There are several reasons for this: oblateness (flattening at the poles), the bulge in the North Pacific, mountains and ocean trenches, etc.

base-ten units and their relative sizes. Furthermore, students need to learn that, as in the example of 'radius of Earth', in the real world, it is rare either to need to know, or even to be able to know, more than the two or three largest base-ten pieces in the number describing some quantity. For most practical purposes, a number is a two- or three-digit number times a (perhaps large) power of 10. This might be a goal for WNA in the twenty-first century.

It is tempting to speculate why the New Math reform of the 1960s did not articulate the process of learning place value as completely as the five-stage description offered above. Perhaps it was a lack of pedagogical insight on the part of participating mathematicians or, more precisely, failure to appreciate the intellectual advances and the years of development needed to progress from one stage to the next. Perhaps it was a residue of the somewhat contemptuous attitude some mathematicians harbour towards base-ten notation, since it involves arbitrary choices and, especially, selection of a base, for which there is no clear mathematical reason. This might explain the introduction of arbitrary bases in the New Math. Perhaps it was because they were still so starry-eyed about the triumph of set theory in establishing foundations for mathematics that mundane classroom issues such as arithmetic did not engage their attention. Whatever the reason(s) behind it, this failure can serve as an exhibit for the claim that mathematical expertise is not the only prerequisite for understanding and positively influencing mathematics education.

6.5 Comments on Particular Sections of Chapter 5[3]

6.5.1 Comments on Section 5.3.1

The linguistic issues of teaching the base-ten place value system are perhaps the feature of WNA that benefits most from cross-national comparisons. In the USA, we have been aware since the paper (Miller and Zhu 1991) pointing out the strict compatibility of Chinese spoken number names with PVN and the comparative disadvantage English speakers have in learning the principles of place value, since they are obscured at the beginning by the irregularity of the -teen numbers and, to a somewhat lesser extent, of the -ty numbers. However, we learn in Sect. 5.3.1.1 that English presents relatively mild problems and that several European languages such as French and Danish are much worse in this regard. My heart goes out to Danish school children trying to make sense of 70 when the name for it is 'three and a half four'!

[3]The section numbers refer to sections of Chap. 5.

6.5.2 Comments on Section 5.3.1.2

In Sect. 5.3.1.2, we learn that children in Algeria have the added burden of trying to translate between several different vernaculars with contradictory conventions, compounded by translational ambiguities.

Perhaps a way to help children get around this kind of linguistic obstacle is to treat the base-ten system for what it is in almost all countries – an imported piece of a foreign language – and to make the translation from traditional number names to 'structural names', or 'mathematics names', that explicitly describe the base-ten structure of each number, a topic of study. This would include explicitly discussing the -teen numbers as being made from one 10 and some 1s and making sure that students could translate between their traditional names and the structural descriptions. Likewise, the -ty numbers (20, 30, …, 90) would be explicitly identified as a certain number of tens and the general two digit numbers as being the sum of some 10s and some 1s. The work of Fuson (e.g. Fuson and Briars (1990), Fuson et al. (1997)) gives some support to this approach.

Beyond helping children to translate between their traditional names and the quantity meanings of numbers with two or three digits, this approach would have the advantage of permitting explicit attention to be paid to the base-ten pieces and how the base-ten structure facilitates computation. General rules could be enunciated that describe what needs to be done to add or to multiply.

> When adding two base 10 numbers, we add the 1s (from the two numbers) together, we add the 10s together, we add the 100s together, etc. Then, if we get more than 10 of any base 10 unit, we convert 10 of it to 1 of the next higher unit.
>
> When multiplying two numbers, we multiply each base 10 piece of one factor, with each base 10 piece of the other factor. Then we sum all these products.
>
> The multiplication of two base 10 pieces amounts to multiplying their digits, and multiplying the base 10 units, and taking the product of these.

These descriptions can be stated briefly or more completely, as appropriate for the context. Besides encapsulating the main principles of base-ten computation in compact form, these general rules have the advantage that they can be taught progressively, starting with two-digit addition and one-digit by two-digit multiplication, and can be formulated more and more generally as students work with larger numbers. The standard column-wise procedures for paper-and-pencil calculation can then be presented as mechanically simple ways to actualise the principles of addition and multiplication formulated as above. This approach would also easily afford discussion of the role of the rules of arithmetic in ensuring that the general recipes formulated above are valid. All this would in turn prepare students for learning the later stages of place value, in the list of five stages given above, and give them a chance of understanding the whole system, which currently seems to be a rare accomplishment.

6.5.3 Comments on Section 5.4.2

The misunderstanding, between the university mathematician Rick and several K-12 mathematics teachers, mentioned by Cooper (2015), makes an interesting study for another mathematician. The miscommunication may have occurred in the interpretation of 'numerical expression'. Rick may have meant this to mean, 'Does this symbol "8(1)" signify a number?', while the teachers may have taken it to mean, 'Is this a well-defined expression involving numbers?' It is the latter, but it is not the former.

The situation is primed for confusion by the frequent use of the word 'division' to signify either division in the usual sense of rational numbers or division with remainder (DWR).

DWR is not an operation on whole numbers in the same sense that addition or multiplication is an operation. That is to say, DWR does not take a pair of whole numbers and return a single whole number: it produces a *pair* of whole numbers, which play very different roles in the process of DWR. One number is the DWR 'quotient' and the other number is the remainder. Considered in this way, DWR defines a somewhat complicated function from pairs of whole numbers to pairs of whole numbers. It is a rather different animal from rational number division, which takes a pair of rational numbers and returns a *single* rational number.

The confusion is further encouraged by the use of the notation 25:3, which is very similar to the usual fraction notation 25/3 (and probably intended to be similar!). The symbol 25/3 denotes the (rational) number x such that $3x = 25$. However, the notation 25:3 stands for the pair of whole numbers q and r, such that $3q + r = 25$, with r understood to satisfy $0 \leq r < 3$. These numbers are of course 8 and 1.

To emphasise the difference between DWR and the operation of division for rational numbers, instead of writing 25:3 = 8(1), which is trying to make DWR look as much as possible like actual rational number division (RND), we might try to emphasise the distinction and try to make DWR look different from RND. To do this, we might define the 'DWR function', which would take a pair of whole numbers (n,d) to the pair of numbers (q,r) such that $n = qd + r$, with the remainder r understood to satisfy the key condition $0 \leq r < d$. Thus, we would write

$$\text{DWR }(25,3) = (8,1)$$

to emphasise the function aspect of DWR. (This notation might not be so student-friendly, however!) The DWR function is not one-to-one (in fact, it is infinite-to-one), is not RND and indeed is not compatible with RND, or with multiplication in the whole numbers, so there should not be any expectation that, just because

$$\text{DWR }(25,3) = (8,1) = \text{DWR }(33,4) = \text{DWR }(41,5) = \text{DWR }(49,6) = \text{DWR }(57,7),$$

etc., that we can conclude that 25/3 = 41/5 = 57/7, etc. But the notation DWR(25,3) = 25:3 tries to make DWR look like RND and sets things up for the confusion that Rick and his teachers experienced.

Besides the definition of the DWR relation as deriving from the equation $n = qd + r$, there is another conventional notation that can adequately express the relationships involved in DWR – the notation of mixed numbers. This would allow us to write

$$\frac{25}{3} = 8\frac{1}{3}$$

Here the conventional interpretation of the right hand side is as a sum:

$$8\frac{1}{3} = 8 + \frac{1}{3}$$

This notation is more or less equivalent to the notation that Rick proposed to the teachers, but avoids the tricky sign, whose interpretation is at the core of the confusion. Understanding that 25:3 = 8(1) means that $\frac{25}{3} = 8\frac{1}{3}$ and that 41:5 = 8(1) means that $\frac{41}{5} = 8\frac{1}{5}$ should help cure people from wanting to conclude equality of the left-hand sides implies equality of the right-hand sides. The main point to be clear on is that, while it might be defensible to call 8(1) a 'numerical expression', it is *not* a number.

6.6 Conclusion

In these comments, I have tried to reinforce the theme of Chap. 5 that WNA is not a simple matter. This can be seen in the historical development of the base-ten place value system, which was at best incompletely realised by early civilisations that began using mathematics heavily. It can also be seen in the conceptual structure, which is at best rather incompletely taught in many countries. Finding better, more effective and more conceptually complete ways of teaching WNA should be a focus of research. Finally, the fact that many issues of current importance (national budgets, climate science, big data) require dealing with large numbers that are known only approximately indicates placing more instructional emphasis on a global understanding of base-ten structure and, in particular, on the base-ten pieces and their relative sizes.

References

Eves, H. (1969). *In mathematical circles*. Boston: Prindle, Weber & Schmidt.
Fuson, K. C., & Briars, D. J. (1990). Using a base-ten blocks learning/teaching approach for first and second grade place-value and multidigit addition and subtraction. *Journal for Research in Mathematics Education, 21*(3), 180–206.
Fuson, K. C., Smith, S. T., & Lo Cicero, A. (1997). Supporting Latino first graders' ten-structured thinking in urban classrooms. *Journal for Research in Mathematics Education, 28*(6), 738–766.
Miller, K. F., & Zhu, J. (1991). The trouble with teens: Accessing the structure of number names. *Journal of Memory and Language, 30*(1), 48–68.
Newman, J. R. (1956). *The world of mathematics*. New York: Simon and Schuster.
Thanheiser, E. (2009). Preservice elementary school teachers' conceptions of multi-digit whole numbers. *Journal for Research in Mathematics Education, 40*(3), 251–281.
Thanheiser, E. (2010). Investigating further preservice teachers' conceptions of multi-digit whole numbers: Refining a framework. *Educational Studies in Mathematics, 75*(3), 241–251.

Cited papers from Sun, X., Kaur, B., & Novotna, J. (Eds.). (2015). Conference proceedings of the ICMI study 23: Primary mathematics study on whole numbers. Retrieved February 10, 2016, from www.umac.mo/fed/ICMI23/doc/Proceedings_ICMI_STUDY_23_final.pdf

Cooper, J. (2015). Combining mathematical and educational perspectives in professional development (pp. 68–75).
Howe, R. (2015). The most important thing for your child to learn about arithmetic (pp. 107–114).