A HISTORY OF ZERO

MARINA TOMA

One of the commonest questions which the readers of this archive ask is: Who discovered zero? Why then have we not written an article on zero as one of the first in the archive? The reason is basically because of the difficulty of answering the question in a satisfactory form. If someone had come up with the concept of zero which everyone then saw as a brilliant innovation to enter mathematics from that time on, the question would have a satisfactory answer even if we did not know which genius invented it. The historical record, however, shows quite a different path towards the concept. Zero makes shadowy appearances only to vanish again almost as if mathematicians were searching for it yet did not recognise its fundamental significance even when they saw it.

The first thing to say about zero is that there are two uses of zero which are both extremely important but are somewhat different. One use is as an empty place indicator in our place-value number system. Hence in a number like 2106 the zero is used so that the positions of the 2 and 1 are correct. Clearly 216 means something quite different. The second use of zero is as a number itself in the form we use it as 0. There are also different aspects of zero within these two uses, namely the concept, the notation, and the name. (Our name "zero" derives ultimately from the Arabic *sifr* which also gives us the word "cipher".)

Neither of the above uses has an easily described history. It just did not happen that someone invented the ideas, and then everyone started to use them. Also it is fair to say that the number zero is far from an intuitive concept. Mathematical problems started as 'real' problems rather than abstract problems. Numbers in early historical times were thought of much more concretely than the abstract concepts which are our numbers today. There are giant mental leaps from 5 horses to 5 "things" and then to the abstract idea of "five". If ancient peoples solved a problem about how many horses a farmer needed then the problem was not going to have 0 or -23 as an answer.

One might think that once a place-value number system came into existence then the 0 as an empty place indicator is a necessary idea, yet the Babylonians had a place-value number system without this feature for over 1000 years. Moreover there is absolutely no evidence that the Babylonians felt that there was any problem with the ambiguity which existed. Remarkably, original texts survive from the era of Babylonian mathematics. The Babylonians wrote on tablets of unbaked clay, using cuneiform writing. The symbols were pressed into soft clay tablets with the slanted edge of a stylus and so had a wedge-shaped appearance (and hence the name cuneiform). Many tablets from around 1700 BC survive and we can read the original texts. Of course their notation for numbers was quite different from ours (and not based on 10 but on 60) but to translate into our notation they would not distinguish between 2106 and 216 (the context would have to show which was intended). It was not until around 400 BC that the Babylonians put two wedge symbols into the place where we would put zero to indicate which was meant, 216 or 21 " 6.

The two wedges were not the only notation used, however, and on a tablet found at Kish, an ancient Mesopotamian city located east of Babylon in what is today south-central Iraq, a different notation is used. This tablet, thought to date from around 700 BC, uses three hooks to denote an empty place in the positional notation. Other tablets dated from around the same time use a single hook for an empty place. There is one common feature to this use of different marks to denote an empty position. This is the fact that it never occured at the end of the digits but always between two digits. So although we find 21 " 6 we never find 216 ". One has to assume that the older feeling that the context was sufficient to indicate which was intended still applied in these cases.

If this reference to context appears silly then it is worth noting that we still use context to interpret numbers today. If I take a bus to a nearby town and ask what the fare is then I know that the answer "It's three fifty" means three pounds fifty pence. Yet if the same answer is given to the question about the cost of a flight from Edinburgh to New York then I know that three hundred and fifty pounds is what is intended.

We can see from this that the early use of zero to denote an empty place is not really the use of zero as a number at all, merely the use of some type of punctuation mark so that the numbers had the correct interpretation.

Now the ancient Greeks began their contributions to mathematics around the time that zero as an empty place indicator was coming into use in Babylonian mathematics. The Greeks however did not adopt a positional number system. It is worth thinking just how significant this fact is. How could the brilliant mathematical advances of the Greeks not see them adopt a number system with all the advantages that the Babylonian place-value system possessed? The real answer to this question is more subtle than the simple answer that we are about to give, but basically the Greek mathematical achievements were based on geometry. Although Euclid's *Elements* contains a book on number theory, it is based on geometry. In other words Greek mathematicians did not need to name their numbers since they worked with numbers as lengths of lines. Numbers which required to be named for records were used by merchants, not mathematicians, and hence no clever notation was needed.

Now there were exceptions to what we have just stated. The exceptions were the mathematicians who were involved in recording astronomical data. Here we find the first use of the symbol which we recognise today as the notation for zero, for Greek astronomers began to use the symbol O. There are many theories why this particular notation was used. Some historians favour the explanation that it is omicron, the first letter of the Greek word for nothing namely "ouden". Neugebauer, however, dismisses this explanation since the Greeks already used omicron as a number - it represented 70 (the Greek number system was based on their alphabet). Other explanations offered include the fact that it stands for "obol", a coin of almost no value, and that it arises when counters were used for counting on a sand board. The suggestion here is that when a counter was removed to leave an empty column it left a depression in the sand which looked like O.

Ptolemy in the *Almagest* written around 130 AD uses the Babylonian sexagesimal system together with the empty place holder O. By this time Ptolemy is using the symbol both between digits and at the end of a number and one might be tempted to believe that at least zero as an empty place holder had firmly arrived. This, however, is far from what happened. Only a few exceptional astronomers used the notation and it would fall out of use several more times before finally establishing itself. The idea of the zero place (certainly not thought of as a number by Ptolemy who still considered it as a sort of punctuation mark) makes its next appearance in Indian mathematics.

The scene now moves to India where it is fair to say the numerals and number system was born which have evolved into the highly sophisticated ones we use today. Of course that is not to say that the Indian system did not owe something to earlier systems and many historians of mathematics believe that the Indian use of zero evolved from its use by Greek astronomers. As well as some historians who seem to want to play down the contribution of the Indians in a most unreasonable way, there are also those who make claims about the Indian invention of zero which seem to go far too far. For example Mukherjee in [6] claims:-

... the mathematical conception of zero ... was also present in the spiritual form from 17 000 *years back in India.*

What is certain is that by around 650AD the use of zero as a number came into Indian mathematics. The Indians also used a place-value system and zero was used to denote an empty place. In fact there is evidence of an empty place holder in positional numbers from as early as 200AD in India but some historians dismiss these as later forgeries. Let us examine this latter use first since it continues the development described above.

In around 500AD Aryabhata devised a number system which has no zero yet was a positional system. He used the word "kha" for position and it would be used later as the name for zero. There is evidence that a dot had been used in earlier Indian manuscripts to denote an empty place in positional notation. It is interesting that the same documents sometimes also used a dot to denote an unknown where we might use x. Later Indian mathematicians had names for zero in positional numbers yet had no symbol for it. The first record of the Indian use of zero which is dated and agreed by all to be genuine was written in 876.

We have an inscription on a stone tablet which contains a date which translates to 876. The inscription concerns the town of Gwalior, 400 km south of Delhi, where they planted a garden 187 by 270 hastas which would produce enough flowers to allow 50 garlands per day to be given to the local temple. Both of the numbers 270 and 50 are denoted almost as they appear today although the 0 is smaller and slightly raised.

We now come to considering the first appearance of zero as a number. Let us first note that it is not in any sense a natural candidate for a number. From early times numbers are words which refer to collections of objects. Certainly the idea of number became more and more abstract and this abstraction then makes possible the consideration of zero and negative numbers which do not arise as properties of collections of objects. Of course the problem which arises when one tries to consider zero and negatives as numbers is how they interact in regard to the operations of arithmetic, addition, subtraction, multiplication and division. In three important books the Indian mathematicians Brahmagupta, Mahavira and Bhaskara tried to answer these questions.

Brahmagupta attempted to give the rules for arithmetic involving zero and negative numbers in the seventh century. He explained that given a number then if you subtract it from itself you obtain zero. He gave the following rules for addition which involve zero:-

The sum of zero and a negative number is negative, the sum of a positive number and zero is positive, the sum of zero and zero is zero.

Subtraction is a little harder:-

A negative number subtracted from zero is positive, a positive number subtracted from zero is negative, zero subtracted from a negative number is negative, zero subtracted from a positive number is positive, zero subtracted from zero is zero.

Brahmagupta then says that any number when multiplied by zero is zero but struggles when it comes to division:-

A positive or negative number when divided by zero is a fraction with the zero as denominator. Zero divided by a negative or positive number is either zero or is expressed as a fraction with zero as numerator and the finite quantity as denominator. Zero divided by zero is zero.

Really Brahmagupta is saying very little when he suggests that n divided by zero is $n/0$. Clearly he is struggling here. He is certainly wrong when he then claims that zero divided by zero is zero. However it is a brilliant attempt from the first person that we know who tried to extend arithmetic to negative numbers and zero.

In 830, around 200 years after Brahmagupta wrote his masterpiece, Mahavira wrote *Ganita Sara Samgraha* which was designed as an updating of Brahmagupta's book. He correctly states that:-

... a number multiplied by zero is zero, and a number remains the same when zero is subtracted from it.

However his attempts to improve on Brahmagupta's statements on dividing by zero seem to lead him into error. He writes:-

A number remains unchanged when divided by zero.

Since this is clearly incorrect my use of the words "seem to lead him into error" might be seen as confusing. The reason for this phrase is that some commentators on Mahavira have tried to find excuses for his incorrect statement.

Bhaskara wrote over 500 years after Brahmagupta. Despite the passage of time he is still struggling to explain division by zero. He writes:-

A quantity divided by zero becomes a fraction the denominator of which is zero. This fraction is termed an infinite quantity. In this quantity consisting of that which has zero for its divisor, there is no alteration, though many may be inserted or extracted; as no change takes place in the infinite and immutable God when worlds are created or destroyed, though numerous orders of beings are absorbed or put forth.

So Bhaskara tried to solve the problem by writing $n/0 = \infty$. At first sight we might be tempted to believe that Bhaskara has it correct, but of course he does not. If this were true then 0 times ∞ must be equal to every number n, so all numbers are equal. The Indian mathematicians could not bring themselves to the point of admitting that one could not divide by zero. Bhaskara did correctly state other properties of zero, however, such as $0^2 = 0$, and $\sqrt{0} = 0$.

Perhaps we should note at this point that there was another civilisation which developed a place-value number system with a zero. This was the Maya people who lived in central America, occupying the area which today is southern Mexico, Guatemala, and northern Belize. This was an old civilisation but flourished particularly between 250 and 900. We know that by 665 they used a place-value number system to base 20 with a symbol for zero. However their use of zero goes back further than this and was in use before they introduced the place-valued number system. This is a remarkable achievement but sadly did not influence other peoples.

The brilliant work of the Indian mathematicians was transmitted to the Islamic and Arabic mathematicians further west. It came at an early stage for al-Khwarizmi wrote *Al'Khwarizmi on the Hindu Art of Reckoning* which describes the Indian place-value system of numerals based on 1, 2, 3, 4, 5, 6, 7, 8, 9, and 0. This work was the first in what is now Iraq to use zero as a place holder in positional base notation. Ibn Ezra, in the 12^{th} century, wrote three treatises on

numbers which helped to bring the Indian symbols and ideas of decimal fractions to the attention of some of the learned people in Europe. *The Book of the Number* describes the decimal system for integers with place values from left to right. In this work ibn Ezra uses zero which he calls galgal (meaning wheel or circle). Slightly later in the 12^{th} century al-Samawal was writing:-
If we subtract a positive number from zero the same negative number remains. ... if we subtract a negative number from zero the same positive number remains.

The Indian ideas spread east to China as well as west to the Islamic countries. In 1247 the Chinese mathematician Ch'in Chiu-Shao wrote *Mathematical treatise in nine sections* which uses the symbol O for zero. A little later, in 1303, Zhu Shijie wrote *Jade mirror of the four elements* which again uses the symbol O for zero.

Fibonacci was one of the main people to bring these new ideas about the number system to Europe. As the authors of [12] write:-
An important link between the Hindu-Arabic number system and the European mathematics is the Italian mathematician Fibonacci.

In *Liber Abaci* he described the nine Indian symbols together with the sign 0 for Europeans in around 1200 but it was not widely used for a long time after that. It is significant that Fibonacci is not bold enough to treat 0 in the same way as the other numbers 1, 2, 3, 4, 5, 6, 7, 8, 9 since he speaks of the "sign" zero while the other symbols he speaks of as numbers. Although clearly bringing the Indian numerals to Europe was of major importance we can see that in his treatment of zero he did not reach the sophistication of the Indians Brahmagupta, Mahavira and Bhaskara nor of the Arabic and Islamic mathematicians such as al-Samawal.

One might have thought that the progress of the number systems in general, and zero in particular, would have been steady from this time on. However, this was far from the case. Cardan solved cubic and quartic equations without using zero. He would have found his work in the 1500's so much easier if he had had a zero but it was not part of his mathematics. By the 1600's zero began to come into widespread use but still only after encountering a lot of resistance.

Of course there are still signs of the problems caused by zero. Recently many people throughout the world celebrated the new millennium on 1 January 2000. Of course they celebrated the passing of only 1999 years since when the calendar was set up no year zero was specified. Although one might forgive the original error, it is a little surprising that most people seemed unable to understand why the third millennium and the 21^{st} century begin on 1 January 2001. Zero is still causing problems!

In the 15^{th}century the extreme limit of any calculus was the million who was soo big! Three hundred years later the astromomists began to use the billion for counting the stars.

One billion is another very big number if it is used for a quantity of apples for example but not large enough if it is used for the number of molecules. For humans we use a billion for example for things like:

- In our brain we have around three billions of neuronal fibres
- If one would live up to the age of 100 years he would be able to count just to 1000000000 without makeing any stop.
- In 55 years one would breathe half of billion times.
- In 33 years, any human being lived around one billion seconds...

References

[1] Calinger, *A conceptual history of mathematics* (Upper Straddle River, N. J., 1999).
[2] G Ifrah, *From one to zero : A universal history of numbers* (New York, 1987).
[3] G Ifrah, *A universal history of numbers : From prehistory to the invention of the computer* (London, 1998).
[4] G G Joseph, *The crest of the peacock* (London, 1991).
[5] R Kaplan, *The nothing that is : a natural history of zero* (London, 1999).
[6] R Mukherjee, *Discovery of zero and its impact on Indian mathematics* (Calcutta, 1991)
[7] S Giuntini, A discussion concerning the nature of zero and the relation between imaginary and real numbers (Italian), *Boll. Storia Sci. Mat.* **4** (1) (1984), 25-63.
[8] R C Gupta, Who invented the zero?, *Ganita-Bharati* **17** (1-4) (1995), 45-61.
[9] P Mäder, "Wie die Puppe ein Adler sein wollte, der Esel ein Löwe, die Äffin eine Königin - so wollte die Null eine Ziffer sein!" Ein Überblick zur Geschichte der Zahl Null, in *Jahrbuch Überblicke Mathematik, 1995* (Braunschweig, 1995), 39-64.
[10] R N Mukherjee, Background to the discovery of the symbol for zero, in Proceedings of the Symposium on the 1500th Birth Anniversary of Aryabhata I, New Delhi, 1976, *Indian J. Hist. Sci.* **12** (2) (1977), 225-231.
[11] K Muroi, The expressions of zero and of squaring in the Babylonian mathematical text VAT 7537, *Historia Sci.* (2) **1** (1) (1991), 59-62.
[12] L Pogliani, M Randic and N Trinajstic, Much ado about nothing - an introductive inquiry about zero, *Internat. J. Math. Ed. Sci. Tech.* **29** (5) (1998),729–744.
[13] S Ursini Legovich, The origin of the zero in Central American civilization. Comparative analysis with the Hindu case (Spanish), *Mat. Enseñanza No.* **13** (1980), 7-20.
[14] M Ja Vygodskii, L'origine du signe de zéro dans la numération babylonienne (Russian), *Istor.-Mat. Issled.* **12** (1959), 393-420.

Department of Mathematics, Valahia University of Târgovişte,
Bd. Unirii 18-24, Târgovişte, Romania
E-mail address: tmmarina@yahoo.com