THE PI CODE

MICHAEL KEITH
Salem, Oregon

Martin Gardner's fictional Doctor Matrix used to say that, properly interpreted, the number pi (the ratio of the circumference of a circle to its diameter, whose decimal expansion begins 3.14159265358979932384626...) contains the entire history of mankind. In this article I give some results of looking at pi in a relatively new way: as an infinite string of letters derived from its expansion in base 26 or base 27.

BASE 26

Base 26 is one of two fairly natural ways of representing numbers as text using a 26-letter alphabet. The number of interest is expressed numerically in base 26, and then the 26 different base-26 digits are identified with letters as 0=A, 1=B, 2=C, ... 25=Z. Here are the first 100 digits of pi expressed in this way:

D.DRSQLOLYRTRODNLHNQTGKUDQGTUIRXNEQBCKBSZIVQQVGDMELMUEXRO
  IQIYALVUZVEBMIJPQQXLKPLRNCFWJPBYMGGOHJMMQISMS...

Lo! At the 6th digit we find a two-letter word (LO), and only a few digits later we find the three-letter ROD embedded in the four-letter TROD. How many other English words can be found if we continue looking?

First, a few pi facts are in order. The digits of pi (in any base) not only go on forever but behave statistically like a sequence of uniform random numbers. (Mathematically proving that this is the case--the "pi is normal" conjecture--is a deep unsolved problem, but numerical analysis of several billion digits suggests that it is true.) Consequently, pi in base 26 emulates the mythical army of typing monkeys spewing out random letters. Among other things, this implies that any text, no matter how long, should eventually appear in the base-26 digits of pi!

We can use the seemingly-random nature of pi's digits to estimate how many words of various lengths we can expect to find in its first million digits (letters). For example, for 4-letter words each group of consecutive 4 letters in pi is equally likely to be one of the 26^4 possible combinations. My dictionary has roughly 5600 4-letter words, so on the average there should be a valid 4-letter word about once every $(26^4)/5600 = 81$ digits.

Here are the corresponding estimates of how many digits we should expect to scan before finding an N-letter word, for N=2 to 10:

2	4	5	1000	8	5,700,000
3	13	6	14,800	9	140,000,000
4	81	7	272,000	10	3,900,000,000

Dividing these numbers into one million gives an estimate of how many N-letter words should be expected in the first million base-26 digits. For N=7, this gives 3.67, and indeed we found three 7-letter words: SUBPLOT at position 115042, CONJURE at 246556, and DEWFALL at 883265. Counts for the other lengths were also as expected. No 8-letter or longer words were found.

The estimates above are for finding any N-letter word; a specific N-letter word should only occur on the average once every 26^N digits. We should expect to need about 2.5×10^{18} letters in order to find the phrase TO BE OR NOT TO BE (without the spaces) once. We can only get as far as TO BE in the first million.

The very first N-letter word in base-26 pi (for each N) is notable. Remarkably, those words from N=1 to N=8 almost make a little poem: O, lo-- / Rod trod steel. / (Oxygen subplot.) These words occur at positions 6, 5, 11, 10, 6570, 11582 and 115042. The only possible contender for an earlier word that we found is the Oxford English Dictionary word HELLY (obs., "pertaining to hell") at position 5458.

That the first 6-letter word is OXYGEN suggests that pi is truly the very stuff of life! Here are all the 6-letter words we found, in order of appearance:

oxygen salify medics pannes cledgy virial revete prinky Libyan thingy ampler upstep rebuts polity teensy hurroo avower corves exarch fogdom cuphea Bogota adhaka sophic Havana rissoa clangs chinol Bakutu uptube granny snudge deific alters desire beggar Uratic wormer macana reflee optics urnism Ovibos potgun amount drover octopi Bisley ancone muring sozzle defied warted whilst livery minter ambury asaron orgies strack geomys zenith aponia retune tunful unfull empery mutate voicer Kubera Alfuro doolie baldie busher camper bullan scroff exceed cheery skiers

We can also look for words that appear as consecutive letters but running backwards. For each N, the first backwards words we found were OR (12), TRY (10), FILM (140), FILMY (140), FLOUTS (6254) and ALPHORN (458071).

The distribution of these, as expected, is similar to the distribution of the forward words. For example, we found three backwards 7-letter words (ALPHORN, FULLEST, HYLIDAE) and no 8-letter ones.

Before venturing into two dimensions, we mention one more recreation involving the linear string of base-26 pi digits. Where, we ask, do the number names ZERO, ONE, TWO, ... first appear?

ZERO 389247   FOUR 11324   SEVEN 786958
ONE 10087   FIVE 64838   NINE 175372
TWO 13463   SIX 14295   TEN 15276

Neither THREE nor EIGHT appear, nor do number names above TEN. Instead of just looking for the first occurrence, we can note a number name each time it appears. Those which appear in the first million digits, in order, are 14261 22662 25212 21122 12166 61166 12192 22126 11221 22666 61162 12261 62612 66629 22161 61220 66662 61220 12266 22156 26226 11226 66112 61266 22261 11212 66116 11166 61211 21722 22652 22166 26221 22022 26116 16662 26266 6. The Beast Number 666 appears frequently in this string.

THE NEXT DIMENSION

We can provide another "degree of freedom" by arranging the base-26 digits of pi in a two-dimensional array. There are many ways to do this (a spiral, a diagonal zigzag filling the quarter-infinite plane, and so on), but for now we just employ one method, which is to fill an infinite vertical strip S units wide, by writing the first S digits in one horizontal row, then the next S digits in the row below that, and so on. We can then select any portion of the array and look for words that occupy consecutive letters and run in any of the eight possible directions (like a word-search puzzle). Perhaps some words will interlock. Perhaps the words will have something in common. Perhaps we will unlock the Pi Code!

Of course, this is the same thing that was done to "discover" the infamous "Bible Code". Since we can choose the letter distance (S) between rows, this gives us many (in our case, a million or so) different ways of looking at the letter string under study, so the possibility of finding "interesting" arrangements of words is considerably increased compared to a one-dimensional search.

On the left side of the next page is a grid we found starting at position 148655 with S = 14061. This contains the words ALPHA (shown in capitals going diagonally, starting in the lower right) connected with OMEGA, with GOD (lower left corner) nearby! On the other side of the coin, consider the grid on the right, starting at position 255717 with S = 13771, which has DEMON and SATAN interlocked, with DEAR on the bottom row. In this case no diagonals are used, which is even more remarkable. Many other words are present in both arrays; we merely noted the ones that seemed to have a common theme.

Words don't even have to be in straight lines, if that fits our purpose. Consider the S = 2736 array in the vicinity of the word CONJURE, one of the three 7-letter words in linear pi. Connected with CONJURE is HOCUS (going vertically) and POCUS (in an L-shape)!

Both truth as well as crackpottery can be found within pi. Remember the bizarre theories of Alfred Lawson? One of the fundamental principles

of Lawsonry was "zig-zag-and-swirl". Turning our attention to position 49287 with S = 49076, we find the lovely little arrangement below.

```
u r r n d a c i v r g c w n p e u b        a c j w c t h v g r o f d k h c
w f p r z c v k m m p u d p w g l y        l c o h i t n r c z t y a r k d
v y u q s v b m u y s n v m r k l i        u l n j t j c w h w z z e k q p
z a k x u g v s e d h m p l l x l d        i b d q u h k e h b d e e d w p
g n j d u v m x w e s y i e g q i z        w f j j k c x u c z S n h c g a
f q o p q t k u s j s r z o k v v d        c x m t c m m m i h A r i c q j
k e n z k z y j n d c f t s r n b r        z w o x r w x z h m T r r e w q
s m i z d h p s i r d u f u w f o I        g k t t k a y a c m A o k d v q
p u f r e c n l f f z f o q l h b j        z j n a u a D E M O N a k i s n
c h n y e a h O M E G A r d r x k p        v z s v y h d i f f b w x b a s
w a e a i p l d f l o a H p i u s q        t a n k g h e h j e j h j y u u
o n i q e n z b r i n d t P v k z h        j x v y i m d h u t q v g j x u
e q g p l s c r v a s g j s L j l v        f y c d q z o u y l k d v j q t
p i e x s z t t z y v j k p u A l s        p i j c m j w z u t w g m t t e
y g n q q e j l e l v k v w o o h m        c c s b v n g a j h q c w x w j
D r q f n g x k b r p l v e x m f f        a h y r D E A R j k x z r u c v
O y e q h z a x v b q r t p k r s c
G a k z d h s j j o q x f m b h e i
                                                   d v j l k a
       v t h h p j H q a t                         r u m n t z
       o w f z z P O b b d                         e m z Z A G
       o s c i b x C v p l                         h S W I R L
       h l C O N J U R E u                         d e n G v h
       b t s j g t S n z v                         i q i t p o
```

We can explore the arts as well as the sciences. At position 505070 with S = 3999, the array at the left appears. We are exhorted to DIG MODAL BEBOP, a popular form of jazz from the 1950s. If we do we'll certainly feel GLAD (start at the G below MODAL and read upwards). One of the giants of modal jazz was Miles Davis, whose initials appear no less than four times in the grid.

```
j g s o t w q q c c d r h k e k        t l y p t s w w i b b b d m O t
g v a x y f l f j l k u a f e l        n b T S U N I L L s x a q H f j
o z t d l g b l M p w i h s D M        u l k R g x k f c d k r O u q u
j n z g c t j D M z s r q d c n        y c k d Y z k u a m a H s i t z
d o e q l j d o r m r w l u z g        h n y O E m c h d f e E m k p c
i l u i z n s q x l s y g p a q        v l x i T y b m q p m R r z b r
x g t z s o k I b z l v b l r i        v s b a c z w u b p d O z m s s
t D M v s d D I G M y c p o i q        z r d z b e j i v z c n p q s h
h p n l j u B E B O P d t h g m        q p I N m m W A T E R x p h w y
j e t c r y q d w D f e n i l y        v d r p x i t v v f t x z l n g
n n e u z i c v e A l j q b l n        o R g d a v p x f s t u v n v x
u x p l z v j l d L i p f v i o        v O i c d l n v j j h c k i t l
j y y t f c y b q g x p d h e p        i F E s h w w s f u c A u z g c
                                       v y B L i m h i c t q a b c m i
                                       p m g o K j j g r q b o u z w k
                                       e r y k z o i k v g w h p g v l
```

Some grids are rich enough to contain entire sentences or poems. The array on the preceding page starts at position 554766 with S = 1058, and is fruitful enough that we can write a complete 5-7-5 haiku using only words found in the grid:

> Sun, elk in water;
> Oho! For her I'll try to
> Be a hero yet.

Another interesting type of grid is illustrated on the left below; it starts at position 65340 with S = 103986. Note the five 4-letter words all running horizontally grouped into a 4x5 rectangle. This is the largest such rectangle we found for any starting position and S less than a million, and it's even more remarkable because the five words have a similar theme (ARIL is a seed covering, and LERP is an edible insect deposit on a plant). We can say "For my MEAL at DAWN, I will LICK LERP from an ARIL."

In the center grid, we exhibit a near-miss 4x4 word square. Although it is easy to find 3x3 word squares, the nearest we could find to a 4x4 square is the following, containing seven of the required eight words.

The 6x6 grid at the right is, perhaps, an indication that it's time to stop this discussion and move on to something else. After all, it's obvious to whom the square speaks, and it clearly spells out the message "U R (see upper right corner) SICK" (bottom row, backwards)!

```
f s z u y x h t p p     o h e h r a l w p o     M I K E u r
d n u e a q o p c i     r z p W I S T k i x     z K i n h l
u e q o m a x x g v     x d r O V E R d m g     u E q c b m
a b w D A W N r a w     f t q R A T E f q y     j I u p r f
i p e M E A L n z m     h z i D U S T y c a     s T b y j f
m f y L E R P r g v     m H h b o h
c t c A R I L g j a     x k c i s k
l e i L I C K q t c
```

BASE 27

Another way to look at the digits of pi is to express it in base 27, with the extra digit assigned to a space, so that we get a series of "words" (strings of letters surrounded by spaces), not just letters. In the May 1993 Word Ways, Lee Sallows suggested that the most natural assignment is 0=space, so that all the letters are assigned non-zero values. (Otherwise, one of the letters--say, A--will have the value zero, which leads to word pairs like AWAKE and WAKE with the same numerical value, even though it seems more natural for them not to.) Given 0=space, the most obvious scheme for the letters is A=1, B=2, and so on.

The beginning of pi in this system is given on the next page. We have divided the lines at word boundaries (i.e., there is a space at the end of each line).

c.cvezcvbmlyzxmswprpiijzhweemupdrxou jhcfmobyhsijlpjsca
zgxlhqunzwkhdfphtstzoprsnu nhawsjlquvbnqpvzqlwwliytpdauuddkzfgmpcu
fnwsavktwroffceijqrhtlvuqlqnox mjrjmq sqmqscvymhqwjrzkwqdathn
fmwfr fzugxgdjsqpk ckjirtxtiq c

It is harder to construct an N-letter word in base 27 than in base 26, because we have to find an (N+2)-character string consisting of the word with a space preceding and following it. If there are $W(N)$ N-letter words in our dictionary, then in D base-27 digits of pi we should expect to find $(W(N) \times D)/27^{(N+2)}$ N-letter words. For D equal to one million, this works out to 152 1-letter words, 291 2-letter words, 94 3-letter words, 14 4-letter words, and one 5-letter word. The actual numbers observed were 137, 244, 83, 10 and 0. The ten 4-letter words are, in order of appearance:

awry full waar buss pupa bale chic kayo kish ruse

WAAR is an alternate form of "ware" in the OED, and a KISH is a wicker basket.

The first 1-letter word, O, is at position 6456. The first 2-letter word, the Greek letter NU, appears at position 10351, followed by US at 10868. The first 3-letter word is a bit of a poser, because we find a great number of obscure specimens before hitting on a common one, WHO, at position 115288. (Earlier possibilities include LIV, DUP, AAM, YAD, DAR and GES.) The first 4-letter word is, as shown above, AWRY. No 5-letter or longer words show up in the first million digits. The word PI itself occurs twice, at positions 212659 and 979046.

The first backwards words of each length are O (6456), TO (696), PUD (41107), and VETO (10354).

PI AS CIPHER TEXT

Another interesting way of looking at base-27 pi is to consider it as a text encoded with a substitution cipher. As with the two-dimensional approach to base 26 pi, this way of looking at the digits allows us to find a lot more syntactically-correct English texts. It might seem that this would produce many long strings of words (after all, there are 26! ways of assigning letters in a substitution cipher), but as we add more words the letter-pattern constraints they induce rapidly curtail the number of solutions.

Here are a few two- and three-word ciphers, with plausible English translations:

57029 rfsyrcllx eugtyocv FEARFULLY MISPROUD
(how we should feel as we contemplate the mysteries of pi?)

155865 dlahfwi dswzavznr WOLFRAM WEANLINGS

(an indication that pi was invented by Stephen Wolfram—or maybe several exceptionally-young members of his company?)

76615 laig fdbizsrqz lvfrixma EYES NUMERATOR HINTEDLY
(what the math student does on seeing the fraction 355/113)

592835 euuplcycw ch SNOOKERED 'EM
(what pi did to everyone who tried to plumb its mysteries)

Some of pi's short ciphertexts have only one solution using un-abridged Merriam-Webster words. Two such are:

edemymksb u rqoqhibut ANACYCLUS, I REDEPOSIT
(the customer addresses the bank teller's plaint)

vtm rrpgegtmt PSI OOGENESIS
(the psychic farmer says he can increase egg production via brain waves)

The longest solvable ciphertext we found had only five words, but none of its solutions are grammatically interesting.

The longest single word with a valid English counterpart is 814790 wpbjngstikmnuydo VENTRICULOGRAPHY, and this is the only 16-letter specimen we found. The 15-letter ones we discovered were:

dermatoglyphics   polydaemonistic   sulphocarbamide   amphiboliferous
hyperglycosuria   unapproximately   encephalometric   membranaceously
swashbucklering   interparoxysmal   phenylcarbimide   prediscountable

DERMATOGLYPHICS is the longest Merriam-Webster isogram (a word having no repeated letters); it was encoded at a total of six different positions. The "words" formed by pi are most likely to be isograms or near-isograms because each letter of the alphabet has the same chance of appearing.

In this article we have just scratched the surface in exploring the digits of pi as text. Many challenges remain, including extending the search past one million letters, searching for text in other languages, and using non-Roman alphabets.