Scan
W. Hellerstein
M. Gardner
and
Scott Kim

Correspondence related to the Linus & Sally sequences 1977
A6345 > A6346
I have been trying to analyze a mathematical entity (a binary sequence) with a surprisingly complex and subtle form, given its simple inductive definition. To date I have found no complete description of that form, but the partial results I have gotten are tenacious in the sense that the structures they give do not apply everywhere, but nonetheless reestablish themselves each time they break down. In fact I strongly suspect that there is no overall structure to the sequence but that the partial structures reestablish themselves constantly, yielding a sequence where form alternates with disorder, and stretches of order are interrupted by segments of chaos, forever. I am writing to you on the advice of Martin Kruskal to find out what references, if any, have been made to this sequence in the mathematical literature, and what results have been gotten. Almost every mathematician I've talked to has heard of this sequence, but can't tell me anything about it, or where they heard it from! Please help.

This sequence, which I call the Linus sequence, begins like this:

```
1110111101011001...
```

It is associated with the Selly sequence, which goes like:

```
011213152132113241124131...
```

The Linus sequence is defined inductively this way: the first Linus digit is (arbitrarily) 1 and each Linus digit must the first is 0 or 1, chosen so that it prevents the occurrence of a repetition, and that the repetition so prevented is the largest such possible. This rule needs clarification; I'll explain it by demonstration. Linus begins with a 1; the second digit is a 0, for otherwise we would have 11, a repetition. To two places Linus is 10; the third place is a 1 (otherwise, Linus would be 1100, repeating 0) so Linus is 101. Now, for the fourth place we have a dilemma; I would give us 1011, repeating 1 at the end, but C would give 1110, repeating 1. Well, 1C is longer than 1, so we avoid it more; to four places it is 1011. To five places it is 10110 (otherwise we'd have 111111, repeating 1) and to six places it is 101100 (otherwise, 101101, repeating 1).
And so on. The Sally sequence is a record of the lengths of the repetitions prevented; as you see above, the second line digit prevented the repetition of 1, of length 1; the third prevents repeating something of length 1; the fourth, something of length 2, the fifth, of length 1, the sixth, length 2, and so on, so to six places Sally is 0C11217 (the 0 denotes the arbitrariness of the initial 1).

Sally, in effect, gives the local structure of the Linus sequence, and when the computer is rolling out reams and reams of digit-filled paper (I have had Linus and Sally computed out to 1,000,000 terms) the Sally sequence gives a far more comprehensible picture of what's going on.

I gave this sequence the name 'Linus' because it is Linus Van Pelt of Schultz's cartoon strip who gave me the idea of the sequence's inductive definition. Linus, you see, was taking a true-false test for which he hadn't studied by trying to outguess the tester's pattern-breaking habits. "True-false tests almost always begin with a 'true'," he pointed out; then comes a false, after that a true, and after that another true "to break the pattern" he deduced. Continuing in this fashion, he completed the test and happily commented that if you're smart, you can take a test without being smart. (Indeed, Linus would've succeeded brilliantly if it hadn't been for one hitch. Later we found out that the test's answers were exactly opposite to the ones Linus gave. "I tried when I should've failed," he complained. Obviously, the test started with a false and continued 1s, Linus reasoned, yielding a photographic negative of his results! This, incidentally, shows how the initial choice is arbitrary and that all that matters is the structure of the sequence, not the values — unless you're taking a test with it!) Linus and the tester agreed to be right in that a sequence like

C111CC1C11C11CC11C111CC11CC111CC11C11

and

C1121711321632171163241132163241214...

seem devoid of structure (except for suspicious hints, such as the absence of 9's from Sally, and the repetition of 1121 and other quartets later on) until you notice that the middle four of each six is either 1CC1 or C11C.

Interested by this severe restraint on a seemingly chaotic sequence, I cast
- scut and found out that the following conjectures held true for as far as I could tell (at the time, 2000 terms):

1) **Quartet structures** - as above; the middle 4 of each 6 are either 1001 or 1110.

2) **Sally restrictions** - sally numbers are only 1, 2, 3, 4, 5, or a multiple of 6; and that the multiples of 6 occur only on the 'doubletes' (i.e. the pairs of digits between adjacent quartets).

3) **Quartet transition rules**; that adjacent quartets are of the same form exactly if the doublet in between them is 01 or 10. Furthermore, each of the four possible doublets has a unique sally quartet associated with the second quartet. Thus we have 100100010, 110100100, 110100100, 110100100 (Actually, 1721 2121 is a minor variant of 6721, which for yields the same transition). Furthermore, each quartet is associated with one of these, so with the transition rules one knows everything, given only the doubletes.

4) **Locality**; that the sally numbers for the doubletes are usually no greater than six, and thus that the sally numbers are restricted to being equal to 1, 2, 3, 4, 5, and 6. Having reduced quartets to transitions from doubletes, I wished to find similar rules for doubletes, but since non-local effects (i.e. sallys like 7C, 24, and 100) tended to upset things, I could see that any results would have limited usefulness. But in any case, what happens when there are no non-local effects? What is Linus's local behavior?

5) **The thirty-cycle**; that given locality, the doubletes have transition rules of their own, and the whole thing goes in a cycle 5 quartet-doublet groups long (i.e. 30 units long).

\[
\begin{array}{ccccccc}
100100010 & 110100100 & 110100100 & 110100100 & 110100100 & 110100100 & 110100100
\end{array}
\]

If nonlocal sally numbers were forbidden, this cycle would continue forever. I know of only one other cycle which is self-consistent in this sense, but it does not arise under its own steams as does the 7C-cycle at Linus's beginning, and it seems to lack a sort of resiliency that the 7C-cycle has. Consider, the 7C-cycle (and any other cycle) cannot repeat itself indefinitely, so nonlocal sallys always
My main question to you, of course, is whether you've heard about this sequence or not. Linus would clearly be an interesting question in finite math and coding and combinatorial theory, so Kruskal tells me that if anyone anywhere has written anything on it, you would have heard about it, and it seems that somebody, somewhere, sometime, has written something on it! Please send me the references and tell me which theorems I've rediscovered and which ones I've missed. Thank you.

Nathaniel Hellerstein

Nathaniel Hellerstein

232 1938 Hall
Princeton University
Princeton, NJ 08540

415 587 2605
Dear Nathaniel Wellerstein: 18 Feb 77

Thanks for telling me about the Linus and Sally sequences. I certainly had not encountered them before, but I delayed replying because I knew I would be at Bell Labs (yesterday) and would see Erdos. Erdos was quite intrigued by it -- a "crazy" procedure he called it -- and said he was unfamiliar with it. Nor did Neil Sloane know about it, or Jon Graham.

I could be wrong, but I suspect that the mathematicians who think they've seen are probably confusing it with the Thue sequence or something similar.

I will write you again if I come across anyone who knows it. Meanwhile, I'll appreciate it if you'll keep me up-to-date on any progress you make in analyzing it. It's something I'd like to put in the column eventually, after the dust settles a bit and we've all checked with more people.

Sloane asked for a copy of your letter, which I'll send him today. This means the sequence may turn up in the next edition of his handbook.

Best,

Martin Gardner
Dear Nathaniel:  

14 Apr 77

Herewith some interesting comments on Linus and Sally from Scott Kim, a brilliant young Korean student at Stanford. Box 8976, Stanford, Ca. 94305, in case you want to exchange ideas with him.

I had mentioned the sequence to Doug Hofstadter, also at Stanford, and Doug has passed it on to Scott. If you do write to Scott, I'd appreciate a carbon so I can keep my file on the two sequences up to date.

Best,

[Handwritten signature]

PS: In reply to Scott's questions, I've sent him a copy of your letter to me.
Doug tells me that you found interesting the idea of doing
mathematics experimentally by computer. The computer does indeed to
restore the fun of guessing to higher mathematics. Gosper is constantly
doing this, hacking away at infinite series on MACSYMA, an all-purpose
MIT calculating program available across the ARPA net, which does such
things as integration, series summations, and substitutions, as well as
more mundane pocket calculations. He will often start off with some
suspicious identity, then randomly attack it with substitutions and
splitting functions. Often the results take several pages to print out
not to mention many minutes of computer time. Results gotten this way
aren't always easy to check theoretically, since the steps are not
always strictly legal. But in the true experimental spirit, he often
confirms identities by simply having each side calculated to some
ridiculous number of decimal places. A vivid metaphorical description
of equally outlandish mathematical improvisation may be found in
Stamislaw Leb's "The Cyberiad" (Avon paperbacks) in the tale called ...
(If you haven't seen this book yet, waste no time in finding a copy--it
is filled with mathematical delights.)

The Linus sequence, which Doug told me about, is particularly
amenable to an experimental treatment. Especially since nothing else
seems to work. First of all, let me see if I have the problem straight.
Define a "pattern" to be the sequence of bits formed by a block of ones
and zeroes repeated twice in a row. Examples: "00", "0101", "101101".
Start with Linus[0]=0. Then the algorithm for generating Linus[n] is as
follows. Temporarily set Linus[n]=0. Scan the previous members of the
Linus sequence for the longest pattern ending with Linus[n]=0.
Similarly, try setting Linus[n]=1 and scan for the longest pattern
ending with Linus[n]=1. Then whichever choice, one or zero, causes the
longer pattern, Linus[n] is defined to be the OPPOSITE choice.

Another idea was framed as an ESP test. A subject would enter the
room, sit down, and attempt to predict a sequence of, say, 100 decimal
digits in turn. Each time a guess would be made, the correct answer
would be read immediately after. Presumably the predictor would start
noticing some sorts of patterns, and attempt to out-guess them.
Whatever the strategy though, the subject's sequence of guesses would
then be used as the target for the next subject. And so on.
Presumably, the sequence would start becoming less and less random in
some way as the testing fed on itself.

A friend of mine recently computed Linus and Sally out to 12000
places. Some statistics.

Generally, Linus likes to happen in blocks of 6, where the center
4 digits are either "0110" or "1001". This first goes wrong in the
2798th place. Immediately after this disturbance, a 5 occurs in Sally
for the first time. By the time Linus settles back down to blocks of 6,
the phase has been shifted 4 places and three more 5s have occurred.
Thereafter, several 5-hiccups occur in intervals of 2938

11612
8674
5736
2738
In each case, a 5 occurs two places later in Sally. It then takes about 151 terms before the series settles back down to its steady state. During steady state, each of the six columns (mod 6) has its own personality. Generally speaking, the center four columns are quite tame. The first column encroaches on the domain of two digit numbers, while the last column goes completely wild.

<table>
<thead>
<tr>
<th>n mod 6</th>
<th>possible values for Sally(n) based on observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1 3 4 6 38</td>
</tr>
<tr>
<td>1</td>
<td>2 3 8</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3</td>
</tr>
<tr>
<td>3</td>
<td>2 3 4</td>
</tr>
<tr>
<td>4</td>
<td>1 4</td>
</tr>
<tr>
<td>5</td>
<td>1 3 6 24 30 108 528 552</td>
</tr>
</tbody>
</table>

Other numbers, such as 5, 7, 18 and 5876 seem only to occur during twilight zones between steady states.

The next massive discontinuity during the transition just after 11612: a 5876 occurs in place 11751. This breaks a pattern which goes all the way back to the beginning—the first time this has happened since the first occurrence of 30. Three places later, the first 7 appears. There doesn’t appear to be any immediate hope of containing Sally. The way things are going, it is possible that all integral lengths occur eventually.

Intuitive appraisal. Linus and Sally are basically trying very hard to be periodic. A glance at Sally will show that large values tend to occur at regular intervals. Of course this can’t go on forever. The regularity is more like that of the prime numbers, which results from the superposition of grids with different periods. Occasionally a new massive value will inject itself into the world, causing hideous new values to mutate in its wake.

Who invented the Linus sequence? How much is known about it? Why did it originate? Lyle Ramshaw (CS dept, Stanford), who ran off the first 12000 places for me, made the observation that Linus is strange because of edge effects— it starts at a particular place. If we start instead with an infinite prehistory of zeroes

\[ \ldots000000 \]

the next digit must surely be a one.

\[ \ldots0000001 \]

Thereafter, each digit is forced to be a zero to avoid patterns of the form 0000...1 0000...1.

\[ \ldots00000000000000... \]
I assume this happens with any periodic prehistory. Which raises the difficult question: is there an infinite prehistory which generates the regular Linus sequence?

Incidentally, it is easy to see that Sally must grow without bound. For if it (she?) did not, then the sequence would have only a finite inheritance, and thus only a finite number of possible subsequences. Subsequently, Sally would be periodic, which would mean that Linus would be periodic, which violates the original intended aperiodicity of Linus. This leaves open such questions as: how fast does Sally grow? and are there always patterns which go all the way back to the beginning?

The Linus sequence is a natural way to generalize the non-repeating (stutter-free, asymmetric) melodies of three or more tones mentioned in your column several years ago. Since it is not possible to avoid immediately repeated blocks in a sequence using only two different symbols (as it is when three symbols are available), the best we can do is to keep the patterns as short as possible. The Linus sequence describes an algorithm for generating a sequence which tries to keep doing this. The problem is to characterize a sequence as generated by a particular method.

Is it possible to produce a better Linus-like sequence by a different method? That is, is there another sequence of zeroes and ones whose Sally sequence grows more slowly than that for the Linus? By trading off the breaking of short patterns for the avoidance of longer patterns, it may be possible to slow the overall growth of Sally. Assuming Linus contains infinitely many patterns which make it all the way back to the beginning, then Sally(n) grows as n/2. A sequence which grows as n/4:

\[1010010001000000001000000000000001...\]

where the number of zeroes in a block grows by powers of two. As already mentioned, we'll never be able to restrict Sally's growth to a constant. How slowly can Sally grow?

An asymmetric sequence by definition allows no patterns. The Linus sequence relaxes this restriction, allowing only those patterns which are inevitable. Going the other way, we can strengthen the definition of "pattern" in several ways. In a "strongly asymmetric sequence" (SAS) we must avoid forming two adjacent blocks of n symbols each which are permutations of one another. The definition of pattern has been generalized to include two adjacent blocks which are identical under permutation. Alternately, we could define a pattern to be three consecutive occurrences of the same block, instead of just 2.

I have included several papers on asymmetric sequences (all started by Erdos in 1961):

On Nonrepetitive Sequences
Entringer and Jackson

Is there a sequence on four symbols in which no two adjacent segments are permutations of one another?
T. C. Brown