Reset reproduction of a 1975 paper presented Saturday, October 25, 1975 at the 727th Meeting of the American Mathematical Society, held at the Massachusetts Institute of Technology in Cambridge, Massachusetts. Abstract published in *Notices of the AMS*, October 1975, A-660, #727-A5.

Non-Repetitive Binary Sequences

Jon Doyle
Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science
Cambridge, Massachusetts 02139

Abstract

An infinite sequence on two symbols is constructed with no three adjacent identical blocks of symbols and no two adjacent identical blocks of four or more symbols, refuting a conjecture of Entringer, Jackson and Schatz. It is further demonstrated that there is no infinite sequence two symbols with no three adjacent identical blocks of symbols and no two adjacent identical blocks of three or more symbols.

Let M be a finite set of symbols (an alphabet), and let $Seq(M)$ denote the set of all finite or infinite sequences of one or more elements of M. The finite sequences in $Seq(M)$ are also called blocks. Let $A = a_1, a_2, a_3, \ldots$ be a sequence in $Seq(M)$ and let $B = b_1, \ldots, b_n$ be a block of length n. Then the sequence A contains an occurrence of the block B if $a_{i+j} = b_j$ for some $i \geq 0$ and all j such that $1 \leq j \leq n$. We say that A contains a k-block of length n if A contains k adjacent occurrences of some block of length n (e.g., 101010 contains a 3-block of length 2.) The block C is an initial (final) segment of the block B if B contains an occurrence of C which contains the first (last) term of B. Finally, $B \sim C$ means that the blocks B and C are identical.

Thue [5 and 6] constructed an infinite sequence on three symbols with no 2-blocks and an infinite sequence on two symbols with no 3-blocks. Erdos [3, p. 240] questioned the existence of an infinite sequence on two symbols with an upper bound on the length of 2-blocks, a property which Thue's examples lack. Entringer, Jackson and Schatz [2] answered this question by constructing an infinite sequence on two symbols with no 2-blocks of length 3 or greater. Their example contains 3-blocks, however, and they conjectured that there exists no infinite sequence on two symbols with no 3-blocks and an upper bound on the length of 2-blocks. Theorem 1 of this paper shows this conjecture to be false by constructing an infinite sequence on two symbols with no 3-blocks and no 2- blocks of length 4 or greater. Theorem 2 shows the upper bound on the length of 2-blocks obtained in Theorem 1 to be the smallest possible.

For the remainder of this paper, the sequences in Seq({0, 1}) will represent all sequences on two symbols. These sequences will be called binary sequences.

Theorem 1 *There exists an infinite binary sequence with no 3-blocks and no 2-blocks of length 4 or greater.*

Proof: Consider the sequences $R_0, R_1, R_2, \ldots$ on the alphabet $\{a, b, c\}$ where $R_0 = a$ and R_{i+1} is obtained from R_i by replacing each symbol in R_i by a block as follows:

$$\begin{aligned} a &\rightarrow abcab \\ b &\rightarrow acabcb \\ c &\rightarrow acbcacb. \end{aligned}$$

Pleasants [4] proved that the infinite sequence R obtained as the limit of $R_0, R_1, R_2, \ldots$ has no 2-blocks. The present proof will use a variant of this sequence. Let $S_0, S_1, S_2, \ldots$ be sequences on $\{a, b, c, d\}$ such that $S_0 = a$ and S_{i+1} is produced by replacing each symbol in S_1 as follows:

$$\begin{aligned} a &\rightarrow abcab \\ b &\rightarrow acabdb \\ c, d &\rightarrow adbcadb. \end{aligned}$$

Note that the sequence S obtained from $S_0, S_1, S_2, \ldots$ is the same as the sequence R except that certain occurrences of the symbol c have been changed to occurrences of the symbol d. It is clear that if S contains a 2-block then so does R; hence S contains no 2-blocks. In fact, S has the following stronger property:

(1) S contains no two adjacent blocks which differ only in c-d mismatches.

Let A, B, C, D and E be binary blocks as follows:

$$\begin{aligned} A &= 10110010100110 \\ B &= 0110100100110 \\ C &= 10100 \\ D &= 10100101 \\ E &= 1011001001. \end{aligned}$$

Let T be the infinite binary sequence obtained by replacing a, b, c and d in S by the blocks AE, BE, CE and DE, respectively. This sequence will be represented in the following discussion as $T = X_1Y_1X_2Y_2X_3Y_3\ldots$ where each X_i is identical to one of A, B, C or D, and each $Y_i \sim E$. We claim that T has no 3-blocks and no 2-blocks of length 4 or greater.

The proof will make use of the following properties of the sequence T. Both are easily verified.

(2) There exist no $i \geq 0$, $n \geq 1$ such that

$$X_{i+1}\ldots Y_{i+n} \sim X_{i+n+1}\ldots Y_{i+2n}.$$

(3) If Z is an occurrence of the block E in T, then $Z = Y_i$ for some i.

To prove the theorem, it suffices to show that T has no 3-blocks of length 3 or less and no 2-blocks of length 4 or more. The reader may observe that T contains no 3-blocks of length 3 or less by examining the blocks EAE, EBE, ECE and EDE, in which any such 3-block must be contained. A further observation left to the reader is that T contains no 2-blocks of the form Q_1Q_2 $(Q_1 \sim Q_2)$ of length 4 or greater such that Q_1 does not contain an occurrence of the block E. In view of property (3) and the method of construction of the sequence S, this may be verified by checking the blocks $EAEBE$, $EAECE$, $EAEDE$, $EBEAE$, $EBECE$, $EBEDE$, $ECEAE$ and $EDEBE$ for such 2-blocks

The remainder of the proof consists of showing that T contains no 2-blocks of the form Q_1Q_2 such that Q_1 contains an occurrence of the block E. The proof of this property is by contradiction. Assume that T has a 2-block of the form Q_1Q_2 $(Q_1 \sim Q_2)$ such that Q_1 contains, for some positive n, exactly n occurrences of the block E. We will show that this assumption implies that T contains a 2-block of the form $X_{i+1}\ldots Y_{i+n} \sim X_{i+n+1}\ldots Y_{i+2n}$ contradicting property (2).

In view of property (3), let i be such that Y_{i+1} is the first occurrence of E in Q_1, so that Q_1 contains the block $Y_{i+1}\ldots Y_{i+n}$. Then there are two cases to consider: either Y_{i+n+1} is contained in Q_2, or it is split between Q_1 and Q_2.

First case: Assume that Y_{i+n+1} is contained in Q_2. There are three subcases: either X_{i+n+1} is contained in Q_1, or it is contained in Q_2, or it is split between Q_1 and Q_2. If X_{i+n+1} is contained in Q_2, then

$$Q_1 = X_{i+1}\ldots Y_{i+n} \text{ and } Q_2 = X_{i+n+1}\ldots Y_{i+2n}$$

contradicting property (2). It follows that T can contain no such 2-block Q_1Q_2. If X_{i+n+1} is contained in Q_1, then

$$Q_1 = Y_{i+1}\ldots X_{i+n+1} \text{ and } Q_2 = Y_{i+n+1}\ldots X_{i+2n+1}$$

(Property (l) of S forbids the possibilities of

$$X_{i+n+1} \sim C \text{ and } X_{i+2n+1} \sim D$$

or vice versa.) This implies that

$$X_{i+2} \dots Y_{i+n+1} \sim X_{i+n+2} \dots Y_{i+2n+1},$$

contradicting (2). Again we conclude that T can contain no such 2-block $Q_1 Q_2$.

Now assume that X_{i+n+1} is split between Q_1 and Q_2. Then there exist an initial segment W of Q_1 and a final segment Z of Q_2 such that

$$Q_1 Q_2 = W Y_{i+1} \dots Y_{i+n} X_{i+n+1} Y_{i+n+1} \dots Y_{i+2n} Z$$

Then either W or Z identifies one of X_{i+1} or X_{i+2n+1} as being identical with X_{i+n+1}, the possibilities of

$$X_{i+n+1} \sim C \text{ and } X_{i+2n+1} \sim D$$

or vice versa being forbidden by (1). Therefore either

$$X_{i+1} \dots Y_{i+n} \sim X_{i+n+1} \dots Y_{i+2n}$$

or

$$X_{i+2} \dots Y_{i+n+1} \sim X_{i+n+2} \dots Y_{i+2n+1},$$

and either of these possibilities contradicts (2). It follows that T can contain no such 2-block $Q_1 Q_2$.

Second case: Assume that Y_{i+n+1} is not contained in Q_2. Then there exist an initial segment W of Q_1 and a final segment Z of Q_2 such that

$$Q_1 Q_2 = W Y_{i+1} \dots Y_{i+n} X_{i+n+1} Y_{i+n+1} X_{i+n+2} Y_{i+n+2} \dots Y_{i+2n+1} Z.$$

Furthermore, there are segments W_0 and Z_0 such that

$$W = W_0 X_{i+1} \text{ with } X_{i+1} \sim X_{i+n+2}$$

and

$$Z = X_{i+2n+2} Z_0 \text{ with } X_{i+2n+2} \sim X_{i+n+1}$$

(The possibilities of

$$X_{i+n+1} \sim C \text{ and } X_{i+2n+2} \sim D$$

or vice versa is forbidden by (1).) But this means that

$$X_{i+1} \dots Y_{i+n+1} \sim X_{i+n+2} \dots Y_{i+2n+2},$$

which contradicts (2). It follows that T can contain no such 2-block $Q_1 Q_2$

This completes the proof. □

Theorem 2 *There is no binary sequence of length 30 or greater with no 3-blocks and no 2-blocks of length 3 or greater.*

Proof: Following Doyle [1], Figure 1 displays a tree of the terms of all binary sequences starting with 0 and with no 3-blocks and no 2-blocks of length 3 or greater. Exchanging the symbols 0 and 1 throughout Figure 1 produces the rest of the binary sequences with these properties. The reader may observe that all alternatives and extensions not shown produce sequences with either 3-blocks or 2-blocks of length 3 or greater. (Note: Most of the repetitions produced by alternatives or extensions not shown are fairly short. The main exceptions occur in the longest paths where 2-blocks of length 14 are produced.) The proof is completed by noting that the longest of the sequences contained in Figure 1 are of length 29. □

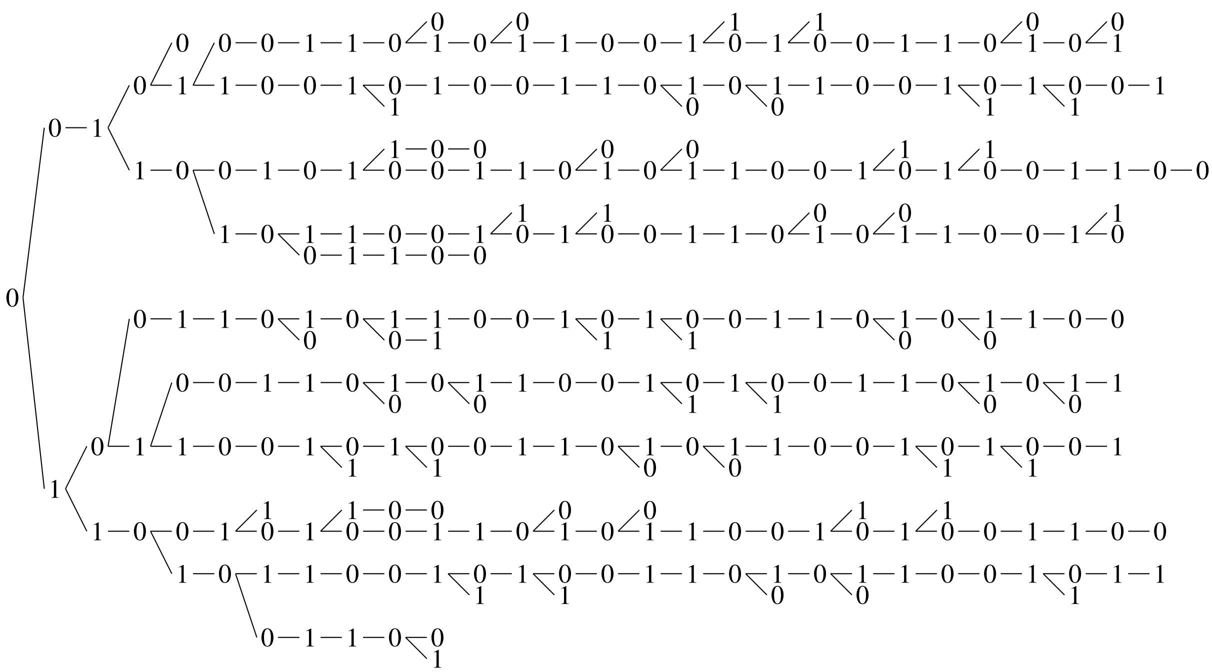

Figure 1: Tree of terms of all binary sequences with no 3-blocks and no 2-blocks of length 3 or greater.

Acknowledgments: The author would like to thank J. A. Schatz for suggesting this problem and for his patience and guidance in this research. Thanks are also due to B. R. Schatz for programming assistance and many valuable criticisms of preliminary versions of this paper, and to R. D. Schafer and L. S. Gray for other assistance in its preparation and presentation. This work was aided by computing facilities made available by the Department of Mathematics of the University of Houston, and support for part of this research was provided by a Fannie and John Hertz Foundation graduate fellowship.

Postscript 2007: This paper was submitted to *Journal of Combinatorial Theory A*, but the reviewer noted that its theorems 1 and 2 were essentially the same as theorems 1 and 2 of another paper the journal had already accepted for publication but not yet published, namely F. M. Dekking, On repetitions of blocks in binary sequences, *J. C. T. A* **20** (1976), 292–299.

References

1. J. Doyle, *Computational investigations of non-repetitive sequences*, Senior Honors Thesis, Department of Mathematics, University of Houston, Houston, Texas, December, 1974.

2. R. C. Entringer, D. E. Jackson and J. A. Schatz, On non-repetitive sequences, *J. Combinatorial Theory, Series A*, **16** (1974), 159–164.

3. P. Erdos, Some unsolved problems, *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, **6** (1961), 221–254.

4. P. A. B. Pleasants, Non-repetitive sequences, *Proc. Cambridge Philos. Soc.*, **68** (1970), 267–274.

5. A. Thue, Über unendliche Zeichenreihen, *Norske Vid. Selsk Skr., I Mat.-Nat. Kl.*, Christiania, 1906, No. 7, 1–22.

6. A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Norske Vid. Selsk Skr., I Mat.-Nat. Kl.*, Christiania, 1912, No. 1, 1–67.