Reprinted from Journal of combinatorial theory
All Rights Reserved by Academic Press, New York and London

Vol. 1, No. 4, December 1966
Printed in Italy

On a Question Concerning Certain Free Submonoids*

M. P. Schützenberger

Faculté des Sciences,
La Sorbonne, Paris, France

Communicated by S. M. Ulam

Abstract

A negative answer is given to a question of Gilbert and Moore concerning the existence of certain type of free submonoids of a free monoid.

Let X^* denote the free monoid (with neutral element e) generated by a fixed set $X \neq \emptyset$ and, for $F \subset X^*$, let F^* denote the submonoid of X^* generated by F. We intend to verify the following property which answers negatively a question raised by Gilbert and Moore in [2, p. 966, line 28] (see [3] for a more complete discussion of this issue).

Property. *Let A be a subset of X^* that satisfies the following three conditions:*

$\mathrm{U}_r(d)$. *There exists a natural number d such that if $a \in A^d$ and $a_1, a_1' \in A$, one has $a_1aX^* \cap a_1'A^* \neq \emptyset$ only if $a_1 = a_1'$.*

$\mathrm{N}''(d)$. *A is maximal among the subsets of X^* that satisfy $U_r(d)$.*

F''. *There is no infinite sequence $a_1, a_2, \ldots, a_n, \ldots$ of elements of A such that each term is a proper left factor of the next one and for each $f \in XX^*$ there exists a natural number m such that $S^mX^* \cap A = \emptyset$ where $S = \{f' \in XX^* : f \in X^*f'\}$.*

* This research was supported in part by the Air Force Office of Scientific Research [AF 61(052)–945].

Then, no $a \in A$ is a proper left factor of another element of A; i.e., A satisfies $\mathbf{U}_r(0)$.

Let $a_1, a_2, ..., a_m, a_1', ..., a'_{m'} \in A$ be such that $a_1a_2 ... a_m = a_1'a_2' ... a'_{m'}$. Multiplying on the right by any $a \in A^d$ and using repeatedly $\mathbf{U}_r(d)$ shows that $a_1 = a_1'$, $a_2 = a_2'$, ..., $m = m'$ and $a_m = a'_{m'}$. Thus $\mathbf{U}_r(d)$ implies that A freely generates A^*; i.e., that A is an "encoding" in the terminology of [2]. For $d = 0$, the condition becomes $a_1X^* \cap a_1'A^* \neq \emptyset$ only if $a_1 = a_1'$ since $A^0 = \{e\}$, as usual. Because $e \in A^*$, it is equivalent to the hypothesis that A has the so called "prefix property," i.e., that for all $a_1, a_1' \in A$ one has $a_1' \in a_1X^*$ only if $a_1 = a_1'$. If there exists a natural number n such that $A \cap X^nX^* = \emptyset$, $\mathbf{U}_r(d)$ becomes equivalent to the "finite delay property" of [2]; if, further, $X = \{x, y\}$, $\mathbf{F}''$ is trivially satisfied and $\mathbf{N}''(d)$ is just another way of expressing the condition (9) of [2], that is,

$$1 = \sum_{m>0} 2^{-m}. \, Card(A \cap X^m).$$

Thus, in this set-up we can state that *no encoding satisfying* (9) *has the finite delay property without having the prefix property.* However, for positive d, the set $A_d = \{x\} \cup (x^dX^* \setminus X^*Xx^dX^*)$ satisfies $\mathbf{U}_r(d)$ and $\mathbf{N}''(d)$ but not $\mathbf{U}_r(d-1)$ nor $\mathbf{F}''$. The set $\{x, xyx\}$ satisfies $\mathbf{U}_r(1)$ and $\mathbf{F}''$ but not $\mathbf{U}_r(0)$ nor $\mathbf{N}''(1)$. The set $\{x, xy, yy\}$ does not satisfy $\mathbf{U}_r(d)$ for any finite d but it satisfies $\mathbf{F}''$ and it is maximal among the sets which freely generate a submonoid of X^*. Finally, letting F denote the set of all left factors of the infinite sequence $xyx^2y^2x^3y^3 ... x^ny^n ...$, the set $FX \setminus F$ satisfies $\mathbf{U}_r(0)$, $\mathbf{N}''(0)$ and $\mathbf{F}''$.

VERIFICATION OF THE PROPERTY. We assume henceforth that A is a non-empty subset of X^* that satisfies $\mathbf{U}_r(d)$, $\mathbf{N}''(d)$, and $\mathbf{F}''$.

REMARK 1. Consider the two sets:

$P =$ the set of all $g \in X^*$ such that for each $f' \in X^*$ one has $A^*gf' \cap A^* \neq \emptyset$ only if $gf' \in A^*$;

$P' =$ the set of all $g' \in X^*$ such that $g'f'X^* \cap A^* \neq \emptyset$ for each $f' \in X^*$.

One has $A^d \subset P = PX^* = P'$.

PROOF. Let $g \in A^d$, $a \in A^*$, and $f' \in X^*$ satisfy $agf' \in A^*$ and show that it implies $gf' \in A^*$. Indeed, a, $agf' \in A^*$ imply $a = a_1a_2 ... a_m$;

$agf' = a_1'a_2' \dots a'_m{}'$ where $a_1a_2, \dots, a_m, a_1', \dots, a'_{m'} \in A$; since $g \in A^d$, the relation $a_1a_2 \dots a_mgf' = a_1'a_2' \dots a'_{m'}$ gives successively $a_1 = a_1'$, $a_2 = a_2', \dots, a_m = a_m'$ by repeated application of $\mathbf{U}_r(d)$. Thus $gf' = a'_{m+1}a'_{m+2} \dots a'_{m'} \in A^*$ and $A^d \subset P$ is proved. The fact that $P = PX^*$ follows instantly from the very definition of P because for $g \in P$, $a \in A^*$, and $f' = f_1f'' \in X^*$ the relations $agf' \in A^*$ and $gf' \in A^*$ are equivalent, respectively, to $ag_1f'' \in A^*$ and $g_1f'' \in A^*$, where $g_1 = gf_1'$.

Keeping the same notations, assume that $P' \neq \emptyset$ and $g' \in P'$. It implies the existence of at least one $g'' \in X^*$ such that $g'g'' \in A^*$ and, for each $f' \in X^*$, of at least one $f'' \in X^*$ such that $g'g''gf'f'' \in A^*$; however, because of $g \in P$ and $g'g'' \in A^*$, this last relation implies $gf'f'' \in A^*$ and we have proved $P \subset P'$ (under the hypothesis $P' \neq \emptyset$). Assume now that $a \in A$ and $f' \in X^*$ satisfy $ag'f' \in A^*$. Because of $g \in P \subset P'$ we can find $h \in X^*$ such that $gh \in A^*$ and, because of $P = PX^*$ we know that $gh \in P$. Because of $g' \in P'$ we can find $h' \in X^*$ such that $g'f'ghh' \in A^*$. Thus $ag'f'ghh' \in A^*$, where $ag'f' \in A^*$ and $gh \in P$, from which we conclude that, in fact, $ghh' \in A^*$. Bringing together these results we see that $A^*g'f' \cap A^* \neq \emptyset$ (since $ag'f' \in A^*$) and $g'f'A^* \cap A^* \neq \emptyset$ (since $g'f'ghh' \in A^*$). Owing to the fact that A freely generates A^*, it is known [1] that these two relations imply $g'f' \in A^*$, and $P' = P$ is proved under the hypothesis $P' \neq \emptyset$.

To end the proof, we assume that A satisfies $\mathbf{U}_r(d)$ only and we show that if $A^d \not\subset P'$ we can find an element $u \in A^dX^*$ such that $B = A \cup \{u\} \neq A$ satisfies $\mathbf{U}_r(d)$ in contradiction with $\mathbf{N}''(d)$.

Let $u \in A^dX^*$ be such that $uX^* \cap A^* = \emptyset$; since $u = au'$ for $a \in A^*$ implies $u'X^* \cap A^* = \emptyset$, we can also assume that $u \not\subset a^{d+1}X^*$. Suppose that b_1, $b_1' \in B$, $b \in B^d$, and $f' \in X^*$ satisfy $b_1bf' \cap b_1'B^* \neq \emptyset$ (i.e., $b_1bf' = b_1'b'$ for some $b' \in B^*$) and show that $b_1 = b_1'$ by considering successively the two cases of $b_1 = u$ and $b_1 \neq u$.

In the first case, $uX^* \cap A^* = \emptyset$ shows that $b_1'b' \in A^*$ is not possible. Let $b' = b_2'b_3' \dots b'_{m'}$ $(b_2', b_3', \dots, b'_{m'} \in B)$ and let j be the least index such that $b_j' = u$. If $j = 1$ we have already the desired conclusion $b_1 = b_1'$. If $j > 1$, the hypothesis $u \in A^dX^*$ implies that $b_1'b_2' \dots b_j'$ has a left factor $a' \in A^{d+1}$; however, a' cannot be a left factor of u since $u \notin A^{d+1}X^*$ nor can it have u as a left factor since $uX^* \cap A^* = \emptyset$. Thus $j > 1$ is impossible and $b_1 = u$ only if $b_1' = b_1$.

In the second case, $b_1 \in A$ and, as above, b_1b has a left factor $a \in A^{d+1}$. Thus $b_1 = b_1'$ follows from $\mathbf{U}_r(d)$ if $b_1'b' \in A^*$. If not, using the same notation and the same reasoning as in the first case, we can exclude

$b_1' = u$ because a cannot have u as a left factor nor be one of its left factors. Finally, for $j > 1$, $b_1'b'$ has a left factor $a' \in A^{d+1}$. Since one of a and a' is a left factor of the other, $b_1 = b_1'$ follows from $\mathbf{U}_r(d)$ and the verification of Remark 1 is concluded.

Remark 2. Let $Q = P \setminus PXX^*$ $(= \{f \in P;\, f \notin PXX^*\})$ and, taking a fixed element $r \in A^d$, let $H' = \{f \in X^* : rf \in A^*\}$ and $H = H' \setminus H'AA^*$. One has $fX^* \cap HQX^* \neq \emptyset$ for each $f \in X^*$ and, for $h, h' \in H$ and $q, q' \in Q$, one has $hqX^* \cap h'q'X^* \neq \emptyset$ only if $h = h'$ and $q = q'$.

Proof. The condition $\mathbf{U}_r(0)$ is equivalent to $P = X^*$, that is, $Q = \{e\}$, and also to $H' = \{e\}$ and, in this case, the remark is trivially true. Thus we can assume that $d > 0$ and $Q \neq \{e\}$.

Let $f \in X^*$. Because of the hypothesis $r \in A^d$ and of $A^d \subset P = P'$ there exists at least one $f' \in X$ such that $rff' \in A^*$ and we can write $ff' = ha'$ with $h \in H'$ and $a' \in A^*$; in fact if $h = h''a''$ where $h'' \in H'$ and $a'' \in A^*$ we have $ff' = h''a''a' \in H'A^*$ and, consequently, we can assume henceforth that $h \in H$. Now, because of $A^d \subset P, ff'r = ha'r = hr'$ where $r' \in P$ and, by the very definition of Q, we have $ff'r = hqf_1$ for some $q \in Q$ and $f_1 \in X^*$. This proves that every $f \in X^*$ is a left factor of at least one element of HQX^*.

Keeping the same notation, assume now that $ff'r$ is equal to $h'q'f_2$ where $h' \in H$, $q' \in Q$, and $f_2 \in X^*$. Without loss of generality we can also assume that h' is a left factor of h, i.e., $h = h'f_3$. By construction, $rhqf_1 = rh'q'f_2 \in A^*$ where, on the one hand, $rh' \in A^*$ because of $h' \in H$ and, on the other, $q' \in P$. Thus, $q'f_2 \in A^*$ and f_3 satisfies $A^*f_3 \cap A^* \neq \emptyset$ (since $rh = rh'f_3$) and $f_3A^* \cap A^* \neq \emptyset$ (since $q'f_2 = f_3qf_1$). As above it implies $f_3 \in A^*$ but, since h and h' belong to $H = H' \setminus H'AA^*$, this is possible only if $f_3 \notin AA^*$, i.e., if $f_3 = e$ and, accordingly, $h = h'$ and $qf_1 = q'f_2$. Since $Q = P \setminus PXX^*$ satisfies $Q \cap QXX^* = \emptyset$, we conclude that $q = q'$, and Remark 2 is verified. In fact, what we have shown is that $H \times Q \rightarrow HQ$ is a bijection and that HQ satisfies $\mathbf{U}_r(0)$ and $\mathrm{N}''(0)$.

We now come to the verification of the property itself. Letting r and H as above, let $h_1, h_2, \ldots, h_n \ldots$ be an infinite sequence of (not necessarily distinct) elements of H such that each term is a left factor of the next one. By the definition of H, there corresponds to each h_n a right factor k_n of r such that $k_nh_n \in A$. Thus only finitely many of the k_n's are different and, because of the first part of $\mathbf{F}''$, we deduce that the same is true for the h_n's. It follows that we can select a fixed $h \in H$ such that $hXX^* \cap H = \emptyset$. Let $S = \{f \in XX^* : h \in X^*f\}$ and $\bar{Q} = \{f \in X^* : fXX^* \cap Q \neq \emptyset\}$. If and only if A satisfies $\mathbf{U}_r(0)$, we have $Q = H = \{e\}$, that is, $S = \bar{Q} = \emptyset$. We show that $\bar{Q} \neq \emptyset$ leads to a contradiction with the second part of $\mathbf{F}''$ by proving first that for any $f \in \bar{Q}$ there exists at least one $s \in S$ such that $sf \in \bar{Q}$.

Indeed, let $f \in \bar{Q}$. Because of $Q = P \setminus PXX^*$ and $P = P'$ we have $f \notin P'$ and, accordingly, there exists at least one $f' \in X^*$ such that $ff'X^* \cap A^* = \emptyset$; since $Q \subset P'$, we have, *a fortiori*, $ff'X^* \cap QX^* = \emptyset$. However, by Remark 2 we know that $hff'X^* \cap HQX^* \neq \emptyset$ and, more accurately that there ex istsone and only one pair $(h', q') \in H \times Q$ such that $hff'X^* \cap h'q'X^* \neq \emptyset$. Because of our choice of h, h' is a left factor of h, i.e. $h = h's$ and $s \neq e$ because otherwise we would have $ff'X^* \cap q'Q \neq \emptyset$ in contradiction with our choice of f'. Thus we have $sff'X^* \cap q'X^* \neq \emptyset$ where $s \in S$. Now q' cannot be a left factor of sf because, taking any $f'' \in X^*$ such that $ff'' \in Q$, it would imply that $hff'' \in HQ$ has a second factor $h'q' \neq hff''$ in HQ. Thus sf is a proper left factor of q' and $sf \in \bar{Q}$ is proved.

Now, since $Q \subset \bar{Q}X^*$, $A^d \subset QX^*$, and $A^d \cap \bar{Q} = \emptyset$, any $f \in \bar{Q}$ is a proper left factor of at least one $a \in A^d$ and we have proved that if $\bar{Q} \neq \emptyset$ one has $A^d \cap S^mX^* \neq \emptyset$ for every natural number m. Recalling that S consists of all the right factors $\neq e$ of h, we see that any factor of any $s \in S^m$ belongs itself to S^*X^*. Since S is a finite set it follows that the relation $A^d \cap S^mX^* \neq \emptyset$ for all m implies that $A \cap S^mX^* \neq \emptyset$ for all m. Since this is excluded by the second part of $\mathbf{F}''$, we must have $\bar{Q} = \emptyset$, that is, $Q = \{e\}$, that is, $P = X^*$, and we have established the conclusion that A satisfies $\mathbf{U}_r(0)$.

OBSERVATION. Using published results on the theory of free monoids one can give to our proposition the following weaker alternative form:

Let A^* be any submonoid of X^* that satisfies the following two conditions:

$\mathbf{U}_d$: $\{f \in X^* : fA^* \cap A^*f \cap A^* \neq \emptyset\} = A^*$.

$\mathbf{N}_d$: $\{f \in X^* : X^*fX^* \cap A^* \neq \emptyset\} = X^*$.

If there exists a natural number n such that

$$X^nX^* \cap (A^* \,/\, (A^* \cap XX^*)^2) \\ = X^nX^* \cap \{f \in X^* \setminus A^* : fA^* \cap X^*f \cap A^* \neq \emptyset\} = \emptyset,$$

then A^* *satisfies*:

$\mathbf{U}_r$: $\{f \in X^* : A^*f \cap A^* \neq \emptyset\} = A^*$ *and*

$\mathbf{N}_r$: $\{f \in X^* : fX^* \cap A^* \neq \emptyset\} = X^*$.

Indeed, $\mathbf{U}_d$ is a necessary and sufficient condition that A^* be freely generated by

$$A = (A^* \cap XX^*)\backslash(A^* \cap XX^*)^2$$

and, if $X^nX^* \cap A = \emptyset$ for large enough n, $\mathbf{N}_d$ is a necessary and sufficient condition that A be a maximal set among the subsets of X^* which freely generate a submonoid of X^* (see [3]). Let us assume

$$\mathbf{U}_d,\ \mathbf{N}_d,\ X^nX^* \cap A = \emptyset, \text{ and } X^nX^* \cap C = \emptyset,$$

where

$$C = \{f \in X^* \,/\, A^* : A^*f \cap fX^* \cap A^* \neq \emptyset\}$$

and let F denote the set of all the left factors of the elements of A^*. Suppose, that $F \cap X^{2n}X^*$ contains a f such that $aff' \in A^*$ and $ff' \notin A^*$ for at least one pair $(a, f') \in A^*xX^*$. Because of $A \cap X^nX^* = \emptyset$ and $f \in X^{2n}X^*$, we can find $g \in X^{n+1}X^*$ and $g' \in X^*$ such that $f = gg'$, $ag = a_1 \in A^*$, and $g'f' \in A^*$. Because of $ff' = gg'f' \in A^*$ we know that $g \notin A^*$ and, because of $g \in F$ we can find $f'' \in X^*$ such that $gf'' = a_2 \in A^*$. Thus $a_2ag = gf''a_1 = a_2a_1 \in A^*$, i.e., $g \in C$ in contradiction with $X^nX^* \cap C = \emptyset$, and we can conclude that $F \cap X^{2n}X^* \subset P$ in the notations of Remark 1. It follows that $A^{2n} \subset P$, and we have proved that A satisfies $\mathbf{U}_r(2n)$. Since A is maximal among the subsets of X^* that freely generate a submonoid of X^*, it is a fortiori maximal among the subsets of X^* that satisfy $\mathbf{U}_r(2n)$ and, since $A \cap X^nX^* = \emptyset$, the condition $\mathbf{F}''$ is trivially verified. Now we can apply our proposition. $\mathbf{U}_r$ is obviously equivalent to $\mathbf{U}_r(0)$ and $\mathbf{N}_r$ is satisfied because it expresses that $P' = X^*$.

References

1. P. M. Cohn, On Subsemigroups of Free Semigroups, *Proc. Amer. Math. Soc.* **13** (1962), 347-351.

2. E. N. Gilbert and E. F. Moore, Variable-Length Binary Encodings, *Bell System Tech. J.* **38** (1959), 933-967.

3. M. Nivat, Théorie générale des Codes, in *Automata Theory* (E. Caianiello, Ed.), Academic Press, New York, 1966.

Printed in Italy by - I.T.E.C. Milan