Algebraic Decision Procedures for Local Testability*

by

ROBERT MCNAUGHTON

Department of Mathematics
Rensselaer Polytechnic Institute
Troy, N. Y.

1. The Notion of Local Testability. A locally testable event is a set of words over an alphabet, membership in which is determined simply by the presence and absence of segments of a certain fixed length. The order in which these segments occur and their frequency are immaterial; however, the left-end segment and the right-end segment are allowed to play special roles. The alphabet Σ is always finite and it turns out that the locally testable events constitute a subfamily of the family of regular events.

Some notation is needed to make this concept precise. Assuming W is a word of length k or more, let $\alpha_k(W)$ be the left-end segment of W of length k; let $\omega_k(W)$ be the right-end length-k segment of W; and let $\beta_k(W)$ be the *set* of interior length-k segments of W. (Thus if W has length k or $k+1$, $\beta_k(W) = \varnothing$, the empty set.) Let Σ^k be the set of all words over Σ of length k, and Σ^* be the set of all words of any length over Σ, including λ, the empty word, which has length zero. Thus $\Sigma^k\Sigma^*$ is the set of all words over Σ whose length is at least k. $|W| = k$ means W has length k.

An event η (i.e., set of words) over a finite alphabet Σ is *k-testable* when, for all words W, $W' \in \Sigma^k\Sigma^*$, $\alpha_k(W) = \alpha_k(W')$, $\beta_k(W) = \beta_k(W')$ and $\omega_k(W) = \omega_k(W')$ imply that either both $W \in \eta$ and $W' \in \eta$, or both $W \notin \eta$ and $W' \notin \eta$. (An event that is k-testable, according to this definition, can be arbitrary about words of length $k-1$ and less.) An event is *locally* testable if it is k-testable, for some k.

Where $|W| \geqq k$, the ordered triple $(\alpha_k(W), \beta_k(W), \omega_k(W))$ will be called the *k-test vector* of W. For a given k and Σ, there are only finitely many k-test vectors, and only finitely many words of length less than k. A k-testable event can be completely characterized by the k-test vectors of words in it of length $\geqq k$, and by the words in it of length $\leqq k+1$. Hence there are only finitely many k-testable events over a given Σ. (For further elucidation, see Chapter 2 of [4].)

It is clear that local testability is a significant notion for the theoretical

* The research for and production of this paper were supported in large part by NSF grant GP-24335.

MATHEMATICAL SYSTEMS THEORY, Vol. 8, No. 1.
© 1974 by Springer-Verlag New York Inc.

treatment of present-day computation, which is so heavily involved with problems of languages and strings of symbols. In the theory of automata, two important results made use of this notion before the notion was clearly formulated. One was a theorem of Medvedev [5] that every regular event is the homomorphic image of a locally testable event. This theorem is easy to prove if one considers how a word is accepted by a finite graph. The second result was the later and deeper theorem by Chomsky and Schützenberger [2] that every context-free language is the homomorphic image of the intersection of a Dyck language and a locally testable event. (A Dyck language, over an alphabet consisting of paired-off letters called inverses, is simply the set of all words that would reduce to the empty word upon successive steps consisting of canceling adjacent occurrences of a letter and its inverse.)

Both of these results used a special kind of locally testable event. Let us say that an event η over Σ is *k-testable in the strict sense* (*k-TSS*) if there exist subsets α, β and ω of Σ^k such that, for all $W \in \Sigma^k\Sigma^*$, $W \in \eta$ if and only if $\alpha_k(W) \in \alpha$, $\beta_k(W) \in \beta$ and $\omega_k(W) \in \omega$. An event is *locally testable in the strict sense* (*LTSS*) if it is *k*-TSS, for some *k*.

An example of an LTSS event is the set of words of $\{a, b\}$ beginning with *aa*, ending with either *aa* or *bb*, but having no interior occurrence of *bb*; here $k = 2$, $\alpha = \{aa\}$, $\beta = \{aa, ab, ba\}$, and $\omega = \{aa, bb\}$. An example of a locally testable event that is not LTSS is the set of words over $\{a, b\}$ that either begin and end with *aa* and have no interior occurrence of *aa*, or begin and end with *bb* with no interior occurrence of *bb*.

Testing for membership in an LTSS event is generally easier than testing for membership in a locally testable event that is not LTSS. Thus, to test a word W for a *k*-TSS event, simply check if $\alpha_k(W) \in \alpha$, $\omega_k(W) \in \omega$, and if each interior segment is in β. Upon the discovery that any length-*k* segment of W fails its test, W can be rejected without further investigation.

Both the Medvedev result and the Chomsky-Schützenberger result use LTSS events: in fact, for both results, a 2-TSS event suffices. It is conceivable that, to many theoreticians, local testability in the strict sense will seem to be a more interesting and natural concept than local testability. For such people, the importance of the latter will consist in the fact that the family of locally testable events is the Boolean closure (i.e., closure under union, intersection and complementation) of the family of LTSS events.

The problem of finding a decision procedure for local testability has been a matter of concern since about 1969. The book [4], in which the concept was introduced, went to press without a solution.

It is tempting to think that a decision procedure similar to the one in [6] based on the reduced state graph should be forthcoming for local testability. However, although there is a fairly straightforward decision procedure for *k*-testability for any fixed *k*, there was no way to prove that a locally testable event whose reduced state graph has *n* states is *k*-testable, where *k* is some computable function of *n*. Of course, such a bound is now known; but it is known only because of work done in establishing a decision procedure which (as it turns out) is much better than the decision procedure that the bound gives us.

(As a byproduct of [1], we can take $n^n + 2$ as a bound. Although this bound is

better than any bound that could come as a byproduct of this paper, the conjecture of Brzozowski and Simon that this bound can be improved upon seems credible.)

The best known decision procedure involves the algebraic concept of what I shall call the *syntactic semigroup* of an event. Both the discovery (or conjecture) of that decision procedure which is given by the Main Theorem of this paper, stated in the next section, and the proof establishing it were achieved independently by two rival research teams: one consisting of myself and Zalcstein, the other consisting of Brzozowski and Simon. Both teams produced both the discovery and the proof during the winter and early spring of 1971 during an interval in which the two teams did not communicate directly with each other. The two proofs use different notions and both will be published. Our proof is contained in this paper. The Brzozowski-Simon proof, to be found in [1], uses a rather sophisticated application of algebraic decomposition theory. Both proofs are lengthy, and there seems no way of substantially simplifying either. It seems justified, therefore, to say that the theorem is a deep one.

I would like to acknowledge my indebtedness to my own team-mate Yechezkel Zalcstein. He contributed a number of significant algebraic thoughts to amplify and clarify some of my own earlier thinking. It was he on our team who conjectured the decision procedure, and much of its justification. It is only a matter of convenience, having to do with work schedules and travel plans, that we decided to publish separately. His contributions to the algebra of local testability and related concepts can be found in [8], [9] and [10].

A decision procedure for local testability in the strict sense seems to have been first discovered by Levy and Freeman [3]. It is fair to say that their problem was not as difficult as finding a decision procedure for local testability. The technique of [6] can, with some effort, be extended to show that an LTSS event η whose reduced state graph has n states is $(n-1)$-TSS. (This bound cannot be improved upon.) Nevertheless, neither the decision procedure based on this fact nor its proof seem as interesting as the algebraic decision procedure presented in Section 5.

2. Technical Introduction. In the literature much use is made of the concept of the syntactic monoid of an event, as the monoid of congruence classes modulo an event of Σ^*. For local testability this concept has a certain failing when the null word is not congruent modulo a locally testable event to any other word. This congruence class consisting of just the null word alone and the resulting monoid identity just get in the way in any useful algebraic application to local testability. (This point was, in retrospect, largely responsible for my own failure to arrive earlier at a decision procedure for local testability.)

To begin with, we shall define an event to be a subset of $\Sigma\Sigma^*$, i.e., the set of all non-null words over the alphabet Σ. Two non-null words W_1 and W_2 are congruent modulo an event η (in symbols $W_1 \equiv W_2 \pmod{\eta}$) if, for all V, $X \in \Sigma^*$, $VW_1X \in \eta$ if and only if $VW_2X \in \eta$. Note that V and X in this definition range of all words of Σ^*, including λ. Although the null word is excluded from membership in an event and from the congruence relation, it will be frequently referred to in this paper.

If α and β are two congruence classes modulo η then $\alpha\beta \subseteq \gamma$, where γ is some congruence class modulo η. (The proofs of the assertions of this paragraph are elementary and well known. See, for example, Chapter 4 of [4]). Thus the congruence classes of an event η form a semigroup under concatenation, which will be called in this paper the *syntactic semigroup*, $SS(\eta)$, of η. $SS(\eta)$ is finite if and only if η is a regular, or finite-state, event. If η is regular and is given as a state graph there is a straightforward algorithm that computes $SS(\eta)$. (For example, one can readily adapt the algorithm for the syntactic monoid given in Chapter 4 of [4].) The mapping ψ: $\Sigma\Sigma^* \to SS(\eta)$ that assigns to each non-null word its congruence class is a homomorphism, to be called the *natural homomorphism* of $\Sigma\Sigma^*$ to $SS(\eta)$.

A finite semigroup has at least one idempotent; that is, an element i such that $i^2 = i$. In fact, for every element a in a semigroup of order n, there exists an $x \leqq n$ such that a^x is an idempotent. (The elementary proof can be found in Chapter 4 of [4]. Furthermore, this proposition is a corollary of Theorem 2.1 below, which is certainly not elementary.)

A semigroup S is quasi-idempotent-commutative (QIC) if, for every x, $y \in S$ and for every idempotent $i \in S$, $(ixi)(ixi) = ixi$ and $(ixi)(iyi) = (iyi)(ixi)$. In other words, S is QIC if and only if $i^2 = i \in S$ implies that iSi is a commutative band (i.e., an idempotent commutative subsemigroup of S). The main theorem of the paper, giving us the best known decision procedure for local testability, can now be stated.

MAIN THEOREM. *A regular event η is locally testable if and only if $SS(\eta)$ is QIC.*

The proof follows from Theorems 3.1 and 4.1 below. (It is fairly straightforward to prove that if η is locally testable then $SS(\eta)$ is QIC. See [8].) Since it is an easy matter to check a finite semigroup to see if it is QIC, the decision procedure is not much more difficult than constructing $SS(\eta)$ from the reduced state graph for η.

We note that the family of all finite QIC semigroups is a pseudo-variety of semigroups. That is to say, (1) a subsemigroup of a finite QIC semigroup is finite and QIC, (2) the homomorphic image of a finite QIC semigroup is finite and QIC and (3) the direct product of two finite QIC semigroups is finite and QIC. (The proof of (1) is obvious, the proof of (3) is straightforward, and the proof of (2) is easy once the following is observed: if ψ is the homomorphism and $\psi(a)$ is an idempotent, then a^n must be an idempotent, for some n, since the semigroup is finite; and $\psi(a^n) = \psi(a)$.) Thus the locally testable events are determined by a pseudo-variety of semigroups. (Cf. Section 5.3 of [4].) This section is closed with the proof of an important proposition for the remaining sections.

THEOREM 2.1 (Ramsey). *For a finite semigroup S there exists a positive integer m_0 (depending only on S) such that for every sequence $t_1, t_2, \cdots, t_m$, where each $t_j \in S$ and $m \geqq m_0$, there exist integers j_1, j_2, j_3, $1 \leqq j_1 < j_2 < j_3 \leqq m$, and an idempotent i of S such that*

$$t_{j_1+1}t_{j_1+2}\cdots t_{j_2} = t_{j_2+1}t_{j_2+2}\cdots t_{j_3} = i.$$

Proof. We apply Theorem B of [7]. Let $\mu = |S|$, $r = 2$ and $n = 3$. For a given m, take $\Gamma_m = \{1, \cdots, m\}$, so that the r-combinations are the sets $\{j, h\}$, $1 \leqq j < h \leqq m$. Let $C_1, \cdots, C_\mu$ be determined by S for a given $t_1, \cdots, t_m$: let $\{j, h\} \in C_k$ if, assuming $j < h$, the product $t_{j+1}t_{j+2}\cdots t_h = s_k$. Then by that theorem of Ramsey, there exists an m_0 such that, for any sequence $t_1, \cdots, t_m$ where $m \geqq m_0$, there is a C_k and a subset $\{j_1, j_2, j_3\}$ of Γ_m such that $\{j_1, j_2\}$, $\{j_1, j_3\}$, $\{j_2, j_3\}$ all belong to C_k. Thus $t_{j_x+1}t_{j_x+2}\cdots t_{j_y} = s_k$, for $x = 1$, $y = 2$, for $x = 1$, $y = 3$, and for $x = 2$, $y = 3$. Since $s_k s_k = t_{j_1+1}\cdots t_{j_2}t_{j_2+1}\cdots t_{j_3} = s_k$, s_k is an idempotent. End of proof.

Note that the m_0 in Ramsey's Theorem B is a function of μ, r and n. In this application r and n are fixed so that m_0 is some function of $\mu = |S|$ alone. It would be interesting to ascertain the minimal combinatorial function f such that for each μ, $m_0 = f(\mu)$ satisfies Theorem 2.1. However, it is not needed for the results of this paper.

Idempotents play an important role in all the algebraic decision procedures of this paper. Theorem 2.1, which could be called "Ramsey's Theorem as applied to finite semigroups", will be an important lemma in many of the proofs.

3. Strings. A *string* in a semigroup is a sequence $a_1, i_1, a_2, i_2, \cdots, a_m, i_m$ such that, for each x, i_x is an idempotent and $a_1a_2\cdots a_x i_x = a_1a_2\cdots a_x$. For each $x \leqq m-1$, the ordered triple (i_x, a_{x+1}, i_{x+1}) is called a *triple of the string*. Two triples (of the same or different strings) (i_x, a_{x+1}, i_{x+1}) and (j_y, b_{y+1}, j_{y+1}) are *equal* if $i_x = j_y$, $a_{x+1} = b_{y+1}$ and $i_{x+1} = j_{y+1}$. $a_1a_2\cdots a_m$ is called the *end product* of the string. Each string has a set of triples associated with it; such sets turn out to be important for the analysis of local testability, as Theorem 3.1 will show. Two strings $a_1, i_1, \cdots, a_m, i_m$ and $b_1, j_1, \cdots, b_n, j_n$ *begin the same* if $a_1 = b_1$ and $i_1 = j_1$. They *end the same* if $i_m = j_n$. We say that a semigroup satisfies the *strings condition* if every two strings that begin the same, end the same, and have the same set of triples also have the same end product.

THEOREM 3.1. *An event η is locally testable if and only if $SS(\eta)$ satisfies the strings condition.*

Proof. Let ψ be the natural homomorphism of $\Sigma\Sigma^*$ to $SS(\eta)$. Suppose first that an event is k-testable and consider two strings $a_1, i_1, \cdots, a_m, i_m$ and $b_1, j_1, \cdots, b_n, j_n$ in which $a_1 = b_1$, $i_1 = j_1$ and $i_m = j_n$, and which have the same set of triples. Let χ be a function that is a subset of the relation ψ^{-1} and such that, for every $s \in SS(\eta)$, $\chi(s) \in \Sigma\Sigma^*$. (In other words, χ is some function such that $\chi\colon S \to \Sigma\Sigma^*$ and $\chi^{-1} \subseteq \psi$.) Let $U_x = \chi(a_x)$ and $V_x = \chi(i_x)$, $1 \leqq x \leqq m$. Let $W_x = \chi(b_x)$ and $X_x = \chi(j_x)$, $1 \leqq x \leqq n$. Note that $U_1 = W_1$, $V_1 = X_1$ and $V_m = X_n$.

Put $Y = U_1V_1^kU_2V_2^k\cdots U_mV_m^k$ and $Z = W_1X_1^kW_2X_2^k\cdots W_nX_n^k$. For any words T and T', TYT' and TZT' have the same k-test vectors, which can be seen as follows. Since $U_1 = W_1$ and $V_1 = X_1$, the two words TYT' and TZT' have the same initial segment of length k; since $V_m = X_n$, they have the same terminal segment of length k. Now let R be any interior segment of TYT' of length k. If R is a segment of $TU_1V_1^k$ then it is also a segment of $TW_1X_1^k$. If it is a segment of V_m^kT' then it is also a segment of X_n^kT'. If R is neither a segment of $TU_1V_1^k$ nor of V_m^kT' then, for some x, it is a segment of $V_x^kU_{x+1}V_{x+1}^k$, since,

for any $V \neq \lambda$, $|V^k| \geqq k$. But, since the two strings have the same set of triples, there is a y such that $i_x = j_y$, $a_{x+1} = b_{y+1}$ and $i_{x+1} = j_{y+1}$. This implies that $V_x = X_y$, $U_{x+1} = W_{y+1}$ and $V_{x+1} = X_{y+1}$. Thus R is a segment of $X_y^k W_{y+1} X_{y+1}^k$. In any case, therefore, if any word R of length k is an interior segment of TYT' then it is an interior segment of TZT'. By symmetric reasoning, if it is an interior segment of TZT' then it is an interior segment of TYT'.

It follows that, since η is k-testable, either both TYT' and TZT' are in η or neither is. Since T and T' are arbitrarily chosen, $\psi(Y) = \psi(Z)$. But these are precisely the end products of the two strings, concluding the proof of Theorem 3.1 in one direction.

The proof the other way requires some other concepts and lemmas. Where $\lambda \neq W = UV$, we need to refer to that point in the word W where U ends and V begins. It is convenient to identify this point in W with the ordered pair (U, V). Let m_0 be determined from $S = SS(\eta)$ as in Theorem 2.1. Where $U \neq \lambda \neq V$, let U' be the length-m_0 terminal segment of U, or U itself if $|U| \leqq m_0$; and let V' be the initial length-m_0 segment of V, or V itself if $|V| \leqq m_0$: then, where i is an idempotent of S, (U, V) is an *i-point* in the word UV if $\psi(U')i = \psi(U')$ and $i\psi(V') = \psi(V')$: (U, V) is an *idempotent point* if it is an i-point for some idempotent i. If (TU, VW) is an idempotent point and $U \neq \lambda \neq V$, then we say that it *occurs inside* the segment UV of $TUVW$.

Now let $k_1, \cdots, k_d$ be all the idempotents of $SS(\eta)$ in arbitrary order. For each idempotent point, (U, V), there is a subset σ of $\{k_1, \cdots, k_d\}$ such that (U, V) is an i-point if and only if $i \in \sigma$. The *principal idempotent* of the idempotent point (U, V) is that member of σ which is earliest in the list $k_1, \cdots, k_d$.

The following observations about this concept will be used in the proofs that follow: (1) The ordered pairs (λ, U) and (U, λ) are not idempotent points, by definition. (2) If (U, V) is an i-point then (TU, VW) is an i-point. (3) If $|U| \geqq m_0$ and $|V| \geqq m_0$, then (U, V) is an i-point if and only if (TU, VW) is an i-point. In this case we shall say that this i-point in TUVW is *determined* by the segment UV. (4) If either $|U| < m_0$ or $|V| < m_0$ then it is possible that (TU, VW) is an i-point but (U, V) is not an i-point. (5) If (U, V) is an idempotent point, $|U| \geqq m_0$ and $|V| \geqq m_0$ then the principal idempotent of (U, V) is the principal idempotent of (TU, VW).

LEMMA. *For any segment of length m_0 or more of any word W, W has an idempotent point that occurs inside that segment.*

Proof. Where $a_1 \cdots a_{m_0}$ is a word of length m_0, apply Theorem 2.1 with $m = m_0$ to the sequence $\psi(a_1), \cdots, \psi(a_{m_0}) = t_1, \cdots, t_{m_0}$. Then, since $t_1 \cdots t_{j_2} i = t_1 \cdots t_{j_2}$ and $it_{j_2+1} \cdots t_{m_0} = t_{j_2+1} \cdots t_{m_0}$, it follows that $(a_1 \cdots a_{j_2}, a_{j_2+1} \cdots a_{m_0})$ is an i-point.

It follows then that every word of length m_0 has an idempotent point. But from this fact the lemma follows by remark (2) above.

COROLLARY. *Each word W of length $4m_0 - 3$ or more has at least two idempotent points (U_1, U_2U_3) and (U_1U_2, U_3), where $W = U_1U_2U_3$, $|U_1| \geqq m_0$, $|U_2| \geqq 1$, and $|U_3| \geqq m_0$.*

Proof. Since m_0 is taken from Theorem 2.1, $m_0 \geqq 2$. Let $W = X_1X_2aX_3X_4X_5$,

where $a \in \Sigma$ and $|X_1| = |X_2| = |X_3| = |X_4| = m_0 - 1$. Then $X_2 a$ has an idempotent point, and aX_3 has an idempotent point. By remark (1) following the definition, these are distinct and satisfy the corollary.

To complete the proof of Theorem 3.1, assume η is not locally testable. Take $k = 4m_0$. Let W and W' be two words of length greater than k and with the same k-test vector, but where $W \in \eta$ and $W' \notin \eta$. Thus $\psi(W) \neq \psi(W')$.

Now if $|W| = k+1$ then its set of interior length-k segments would be empty. The only word having the k-test vector of W would be W itself. Thus since $W \neq W'$, $|W| \geq k+2$ and $|W'| \geq k+2$.

Suppose there are n idempotent points of W and p idempotent points of W'. Let $W = U_1 \cdots U_{n+1}$ where the xth idempotent point of W is $(U_1 U_2 \cdots U_x, U_{x+1} \cdots U_{n+1})$ with principal idempotent i_x, $1 \leq x \leq n$. Let $W' = U_1' \cdots U_{p+1}'$ where the yth idempotent point of W' is $(U_1' \cdots U_y', U_{y+1}' \cdots U_{p+1}')$ with principal idempotent j_y, $1 \leq y \leq p$. Put $a_x = \psi(U_x)$ and $b_y = \psi(U_y')$.

Now the strings $a_1, i_1, \cdots, a_n, i_n$ and $b_1, j_1, \cdots, b_p, j_p$ begin the same, which is proved as follows. Since $k = 4m_0$, the first idempotent point in W occurs inside $\alpha_k(W)$ and is determined by $\alpha_k(W)$. Since $\alpha_k(W) = \alpha_k(W')$, this idempotent point will correspond exactly to the first idempotent point of W' which will have the same principal idempotent by remark (5) following the definitions. Similarly, using the fact that $\omega_k(W) = \omega_k(W')$, the strings end the same and $\psi(U_{n+1}) = \psi(U_{p+1}')$. Since $a_1 \cdots a_n \psi(U_{n+1}) = \psi(W) \neq \psi(W') = b_1 \cdots b_p \psi(U_{p+1}')$, we see that $a_1 \cdots a_n \neq b_1 \cdots b_p$; i.e., the end products of the strings are unequal.

To complete the proof, we must prove that the sets of triples are the same. To this end, we prove that every triple of the first string is a triple of the second string, the converse following by symmetry. Thus consider the triple (i_x, a_{x+1}, i_{x+1}).

Case I. $|U_1 \cdots U_x| \leq m_0$. Then, by the Lemma, $|U_1 \cdots U_{x+1}| \leq 2m_0 < k - m_0$. Put $\alpha_k(W) = U_1 \cdots U_{x+1} Q$. Since $|Q| > m_0$, we find that the first $x+1$ idempotent points of W are determined by the segment $\alpha_k(W)$. Since $\alpha_k(W') = \alpha_k(W)$, $i_1, \cdots, i_{x+1}$ being the principal idempotents of these must be identical, respectively, to $j_1, \cdots, j_{x+1}$. Also $U_1 = U_1', \cdots, U_{x+1} = U_{x+1}'$; thus $a_1 = b_1$, $\cdots$, $a_{x+1} = b_{x+1}$. We therefore have $(i_x, a_{x+1}, i_{x+1}) = (j_x, b_{x+1}, j_{x+1})$.

Case II. $|U_{x+2} \cdots U_{n+1}| \leq m_0$. Then since $\omega_k(W) = \omega_k(W')$, we can reason as in Case I to prove that $|U_{x+1} \cdots U_{n+1}| < k - m_0$ and that $U_{x+1} = U_{x+(p-n)+1}$, $U_{x+2} = U_{x+(p-n)+2}, \cdots, U_{n+1} = U_{p+1}$. We can continue this reasoning to conclude that $(i_x, a_{x+1}, i_{x+1}) = (j_{x+(p-n)}, b_{x+(p-n)+1}, j_{x+(p-n)+1})$.

Case III. $|U_1 \cdots U_x| > m_0$ and $|U_{x+2} \cdots U_{n+1}| > m_0$. Since $|W| \geq k+2$ and $|U_{x+1}| < m_0$ (by the Lemma), we can find X_1, X_2, X_3, X_4 such that $U_1 \cdots U_x = X_1 X_2$, $U_{x+2} \cdots U_{n+1} = X_3 X_4$, $|X_2| \geq m_0$, $|X_3| \geq m_0$, $|X_2 U_{x+1} X_3| = k$, $|X_1| \geq 1$ and $|X_4| \geq 1$. We have $(X_2, U_{x+1} X_3)$ and $(X_2 U_{x+1}, X_3)$ as idempotent points whose principal idempotents are i_x and i_{x+1}, respectively. Also these are adjacent idempotent points in $X_2 U_{x+1} X_3$ in the sense that if $U_{x+1} = RS$, $R \neq \lambda \neq S$, then $(X_2 R, S X_3)$ is not an idempotent point. (For if it were, then $(X_1 X_2 R, S X_3 X_4)$ would be an idempotent point, contradicting the stipulation about the idempotent points of W.)

The word $X_2 U_{x+1} X_3$ must be an interior segment of W', since W and W'

have the same k-test vector. Let $W' = Y_1X_2U_{x+1}X_3Y_4$. If $U_{x+1} = RS$, $R \neq \lambda \neq S$, then (Y_1X_2R, SX_3Y_4) is not an idempotent point of W'. (For otherwise (X_2R, SX_3) would be an idempotent point.) Also the principal idempotents of $(Y_1X_2, U_{x+1}X_3Y_4)$ and $(Y_1X_2U_{x+1}, X_3Y_4)$ are the same, respectively, as the principal idempotents of $(X_2, U_{x+1}X_3)$ and (X_2U_{x+1}, X_3), i.e., i_x and i_{x+1}. For some y, therefore, $Y_1X_2 = U_1' \cdots U_y'$, $U_{x+1} = Y_{y+1}'$, $i_x = j_y$, $i_{x+1} = j_{y+1}$ and $a_{x+1} = b_{y+1}$. Thus $(j_y, b_{y+1}, j_{y+1}) = (i_x, a_{x+1}, i_{x+1})$. This completes the proof of Theorem 3.1.

We now show that Theorem 3.1 gives us a decision procedure of sorts for local testability. We say that in the string $a_1, i_1, \cdots, a_m, i_m$ the subsequence $a_x, i_x, \cdots, a_y, i_y$ is a *redundancy* if $a_1 \cdots a_x = a_1 \cdots a_y$, $i_x = i_y$ and for each z, $x \leqq z \leqq y-1$, the triple $(i_z, a_{z+1}, i_{z+1}) = (i_w, a_{w+1}, i_{w+1})$, where either $w \leqq x-1$ or $w \geqq y$. In this case, note that the string $a_1, i_1, \cdots, a_x, i_y, a_{y+1}, i_{y+1}, \cdots, a_m$ has the same end product, the same beginning, the same ending and the same set of triples as the original string. So, in testing for local testability, we need only examine strings without redundancies.

But if $m \geqq sd(t+2)$ then the string $a_1, i_1, \cdots, a_m, i_m$ has a redundancy, where $s = |SS(\eta)|$, d is the number of idempotents in $SS(\eta)$, and t is the number of distinct triples of $SS(\eta)$. To see this, note that, for some $c \in SS(\eta)$ and $x_1 < x_2 < \cdots < x_{t+2}$, we have $c = a_1 \cdots a_{x_1} = a_1 \cdots a_{x_2} = \cdots = a_1 \cdots a_{x_{t+2}}$, and $i_{x_1} = i_{x_2} = \cdots = i_{x_{t+2}}$. Since there are only t triples, $a_{x_k}, i_{x_k}, \cdots, a_{x_{k+1}}, i_{x_{k+1}}$ must be a redundancy for some k, $1 \leqq k \leqq t+1$.

Thus, in checking for local testability, we need only examine strings $a_1, i_1, \cdots, a_m, i_m$, where $m < sd(t+2)$, of which there are only finitely many. Hence we have a decision procedure for local testability. Admittedly; it is rather clumsy and is not as efficient or interesting as the one given by the Main Theorem.

In [8], Zalcstein introduces the concept of a "locally testable semigroup". If we consider words over the alphabet which is the semigroup itself, then such a word determines a product. A semigroup is *locally testable* if there is a k such that whenever two words have the same k-test vector their products are equal. Zalcstein proves that an event is locally testable if and only if its syntactic semigroup is locally testable.

The interesting thing for this paper is that Theorem 3.1 could have been revised to read, "A semigroup is locally testable if and only if it satisfies the strings condition". Then only slight revisions in the present proof would be needed for a correct proof of the revised Theorem 3.1. For that matter, the main theorem of this paper could have been revised to talk about semigroups instead of events: "A semigroup is locally testable if and only if it is QIC". This whole paper could in this way have been revised to be completely outside of automata theory. In my opinion, however, the concept of a *locally testable event* is more interesting and significant than the concept of a *locally testable semigroup*, which is why the paper is in its present form.

We close this section with a tangential remark on generalized definite events. Such an event η has the property that, for some k and for all words W and W' whose lengths exceed k, $\alpha_k(W) = \alpha_k(W')$ and $\omega_k(W) = \omega_k(W')$ imply that $W \in \eta$ if and only if $W' \in \eta$. The proof of Theorem 3.1 can be simplified to become a proof that an event η is generalized definite if and only if $SS(\eta)$

satisfies the strengthened strings condition. This condition is that all strings S, S' that begin the same and end the same have the same end product. (The strengthening consists of omitting reference to the sets of triples.)

The strengthened string condition is easily proved equivalent to the following: for all a, $b \in SS(\eta)$ and for all idempotent i and j in $SS(\eta)$, $iaj = ibj$. It is interesting to compare this last condition for generalized definiteness with the more interesting result of Zalcstein [8], namely that η is generalized definite if and only if for all $a \in SS(\eta)$ and for all idempotents i, $iai = i$. This latter condition relates to the former condition (although the proof of equivalence is elementary) as the QIC condition relates to the strings condition.

4. QIC Semigroups and the Strings Condition.

THEOREM 4.1. *A semigroup is QIC if and only if it satisfies the strings condition.*

The proof of this theorem, which together with Theorem 3.1, gives us the Main Theorem, occupies this entire section. We begin by noting that a semigroup is QIC if and only if it satisfies the two rules

Q1 $$iai = iaiai,$$

and

Q2 $$iaibi = ibiai,$$

for all idempotents i and for all elements a and b. These rules will be referred to explicitly in the proof below.

It is easy to see that the strings condition formally implies that a semigroup is QIC. Using the strings i, i, ai, i and i, i, ai, i, ai, i, we get Q1. Using the strings i, i, ai, i, bi, i and i, i, bi, i, ai, i, we get Q2. Thus we shall be concerned, for the remainder of this section, with the proof the other way. So we shall assume that a semigroup S is QIC and seek to prove that S satisfies the strings condition.

Where $S = a_1, i_1, \cdots, a_m, i_m$, a string, let $\alpha(S) = (a_1, i_1)$, $\omega(S) = i_m$, $\tau(S) = \{(i_1, a_2, i_2), \cdots, (i_{m-1}, a_m, i_m)\}$, and $\pi(S) = a_1 \cdots a_m$. These are, in words, the beginning, ending, set of triples and the end product of S, respectively.

A *pseudostring* is a sequence that is either a string or becomes a string when an element of the semigroup is attached to the beginning or an idempotent is attached at the end, or both. Thus if $a_1, i_1, \cdots, a_m, i_m$ is a string, then it is a pseudostring and so are the sequence $i_1, \cdots, a_m, i_m$, the sequence $a_1, i_1, \cdots, a_m$, and the sequence $i_1, \cdots, a_m$.

We need a term for a pseudostring that begins and ends with the same idempotent. Borrowing a word from music theory, we call $S = i_1, a_2, i_2, \cdots, a_m, i_m$ *a pseudostring with tonic* i if $i = i_1 = i_m$. We put $\alpha(S) = \alpha(i, S)$, $\omega(S) = \omega(i, S)$, $\tau(S) = \tau(i, S)$ and $\pi(S) = \pi(i, S)$. Our preliminary objective is to prove Lemma 3, which is Theorem 4.1 (in the difficult direction) confined to pseudostrings with a tonic.

Recalling the terminology of sequences, the sequence $S_1 = f_1, \cdots, f_m$ is a subsequence of $S_2 = g_1, \cdots, g_n$ if $m \leqq n$ and if for some $x \leqq n-m$ and for

all $y, f_y = g_{x+y}$. It is an *initial subsequence* if $x = 0$; it is a terminal subsequence if $x = n-m$. Thus we may speak of initial subpseudostrings, and terminal subpseudostrings. We write S_1,S_2 (with a comma but no space) to mean the sequence $f_1,\cdots,f_m, g_1,\cdots,g_n$. If S_1 and S_2 are both strings then so is S_1,S_2. If S_1 and S_2 are both pseudostrings, then S_1,S_2 need not be. But where S_1 and S_2 are pseudostrings ending and beginning, respectively, with idempotent i, then we simply write S_1,S_2 to mean the pseudostring S_1,S_2', where S_2' is S_2 minus its initial i; thus S_1,S_2 is also a pseudostring.

Lemma 1 below is by far the most difficult part of the proof of Theorem 4.1 to read. The reader can help motivate himself by reading Lemma 3 and its short proof that uses Lemma 2. The difficult work in proving Lemma 2 is in the proof of Lemma 1.

For an example, let $S = i, a, j_1, b, j_2, c, i, d, j_2, e, j_1, f, i$, and $S' = i, d, j_2, c, i$. Clearly, $\tau(S') \subseteq \tau(S)$, or, in other terms, $\tau(S, S') = \tau(S)$. We wish to prove $\pi(S, S') = \pi(S)$. The first step in the method illustrating the general proof is to notice that we can get $\pi(S) = \pi(S, d, j_2, e, j_1, f, i)$ by Q1 on i. By Q1 on j_2 we see that the right side equals

$$\pi(S, d, j_2, c, i, d, j_2, e, j_1, f, i, d, j_2, e, j_1, f, i)$$

or, simplifying inside, $\pi(S, S', T_1)$, where T_1 is another pseudostring with tonic i. The problem remains to get rid of T_1 here, but meanwhile let us generalize: in this manner, given S and S' satisfying the hypothesis of Lemma 2 we can always get $\pi(S) = \pi(S, S', T_1)$ for some T_1; the number of steps in the process will increase with the length of S'. This result is almost Lemma 1; it must be strengthened to make T_1 a terminal subpseudostring of S to be applicable in the proof of Lemma 2. In the above example, if we take $T = i, d, j_2, e, j_1, f, i$, we find that T is a terminal subpseudostring of S and that $T_1 = T, T$. Since $\pi(S, S', T_1) = \pi(S, S', T)$ by Q1 on i, we get $\pi(S) = \pi(S, S', T)$, illustrating Lemma 1.

LEMMA 1. *Suppose that S and S' are both pseudostrings with tonic i, and that $\tau(S') \subseteq \tau(S)$. Then there is a terminal subpseudostring T of S with tonic i such that $\tau(S, S', T) = \tau(S)$ and $\pi(S, S', T) = \pi(S)$.*

Proof. Where $S' = i, b_1, j_1, \cdots, j_{n-2}, b_{n-1}, j_{n-1}, b_n, i$, put $S_r = i, b_1, \cdots, b_r, j_r$, for $0 \leq r \leq n$. Assume that $j_0 = j_n = i$ and thus $S_0 = i$ and $S_n = S'$. Noting that (j_{r-1}, b_r, j_r) is a triple of S, take V_r and T_r so that $S = V_r, j_{r-1}, b_r, j_r, T_r$, for $r \geq 1$; take V_0 as the null pseudostring and T_0 so that $S = i, T_0$.

Clearly, $\tau(S, S_r, T_r) = \tau(S)$ for all r. We prove by induction on r that $\pi(S, S_r, T_r) = \pi(S)$. Then, since $S' = S_n$, $T = i, T_n$ will satisfy the Lemma.

Since $S, S_0, T_0 = S, S$, we have $\pi(S, S_0, T_0) = \pi(S)$, by Q1 on i. Now assume as an inductive hypothesis that $\pi(S, S_r, T_r) = \pi(S)$. Put $Q = S, S_{r+1}, T_{r+1}, S_r, T_r$ and $R = S, S_r, T_r, S_{r+1}, T_{r+1}$. Since $S, S_r, T_r = V_{r+1}, j_r, b_{r+1}, j_{r+1}, T_{r+1}, S_r, T_r$, we have

$$\begin{aligned}\pi(S, S_r, T_r) &= \pi(V_{r+1}, j_r, b_{r+1}, j_{r+1}, T_{r+1}, S_r, T_r)\\ &= \pi(V_{r+1}, j_r, b_{r+1}, j_{r+1}, T_{r+1}, S_r, b_{r+1}, j_{r+1}, T_{r+1}, S_r, T_r),\end{aligned}$$

the last equation following by Q1 on j_r. But this last string is precisely Q. Hence, by the inductive hypothesis, $\pi(S) = \pi(Q)$.

Now $\pi(Q) = \pi(R)$, by Q2 on i. And, since $\pi(S, S_r, T_r) = \pi(S)$, we have $\pi(R) = \pi(S, S_{r+1}, T_{r+1})$, by the definition of π, looking at the definition of R. In summary, $\pi(S) = \pi(Q) = \pi(R) = \pi(S, S_{r+1}, T_{r+1})$, completing the inductive step.

LEMMA 2. *If S and S' are pseudostrings with tonic i and if $\tau(S, S') = \tau(S)$, then $\pi(S, S') = \pi(S)$.*

Proof. By Lemma 1, we can find a terminal subpseudostring T of S, with tonic i, such that $\pi(S) = \pi(S, S', T)$. But $\pi(S, S', T) = \pi(S, T, S')$, by Q2 on i, $= \pi(S, S')$, by Q1 on i (since T is a terminal subpseudostring of S with tonic i).

LEMMA 3. *If two pseudostrings with the same tonic i have the same set of triples, then they have the same end products.*

Proof. Let S_1 and S_2 be these pseudostrings. Then, using Lemma 2 and Q2, $\pi(S_1) = \pi(S_1, S_2) = \pi(S_2, S_1) = \pi(S_2)$.

Note that we have achieved our intended objective for the special case where the strings in question have their first idempotent equal to their last.

An idempotent *i occurs to the right of* an idempotent j in a string $S = a_1, i_1, \cdots, a_m, i_m$ if there exist x and y, where $1 \leqq x < y \leqq m$, $i = i_y$ and $j = i_x$. Note that it is possible both for i to occur to the right of j and for j to occur to the right of i. i is *right connected* to j in S if there is a sequence of idempotents $k_1, \cdots, k_p$, $p \geqq 2$, such that $i = k_1$, $j = k_p$ and, for each $x \leqq p-1$, k_x occurs to the right of k_{x+1} in S. If p is the smallest such integer for which such a sequence exists, then p is the *degree* of the connection. Of interest are *connected strings*, i.e., strings in which every idempotent is right connected to every other idempotent. It is easy to see that the string $a_1, i_1, \cdots, a_m, i_m$ is connected if and only if i_1 is right connected to i_m.

LEMMA 4. *Let $S = a_1, i_1, \cdots, a_m, i_m$ be a connected string. Then there exists a string S' in which i_1 occurs to the right of i_m in S', and such that $\alpha(S') = \alpha(S)$, $\omega(S') = \omega(S)$, $\tau(S') = \tau(S)$ and $\pi(S') = \pi(S)$.*

Proof. We prove instead a proposition that clearly implies Lemma 4 by mathematical induction: that, for every such S in which the degree of the right connection between i_1 and i_m is $p \geqq 3$, there is an S'' such that $\alpha(S'') = \alpha(S)$, $\omega(S'') = \omega(S)$, $\tau(S'') = \tau(S)$ and $\pi(S'') = \pi(S)$, but such that the degree of the right connection in S'' between i_1 and i_m is at most $p-1$.

Let $k_1, \cdots, k_p$ be the sequence of idempotents such that $i_1 = k_1$, $i_m = k_p$ and, for each $x \leqq p-1$, k_x occurs to the right of k_{x+1} in S. Note that, since p is the degree of the connection, k_x does not occur to the right of k_{x+2} for any x. In particular, k_1 does not occur to the right of k_3. Thus, for some T_1, T_2, T_3, T_4, T_5, we have $S = a_1, k_1, T_1, k_2, U, T_5$, where $U = T_2, k_1, T_3, k_3, T_4, k_2$. Take $S'' = a_1, k_1, T_1, k_2, U, U, T_5$, which satisfies the requirement. The degree of connection is one less than in S because k_2 is no longer needed, k_1 now occurring to the right of k_3. $\pi(S) = \pi(S'')$ by Q1 on k_2. This concludes the proof.

LEMMA 5. *Let $S = a_1, i_1, \cdots, a_m, i_m$ be a connected string, where $i_1 \neq i_m$. Then there exists a string $S'' = a_1, i_1, S_1, i_m, S_2, i_1, S_3, i_m$, such that $\tau(S_2) = \tau(S'') = \tau(S)$ and $\pi(S'') = \pi(S)$.*

Proof. Let S' be the string that is declared to exist by Lemma 4. Suppose $S' = a_1, i_1, T_1, i_m, T_2, i_1, T_3, i_m$. Take $S_1 = T_1$, take $S_3 = T_3$ and take $S_2 = T_2, i_1, T_1, i_m, T_2, i_1, T_3, i_m, T_2, i_1$. Then $\tau(S_2) = \tau(S') = \tau(S)$. Take $S'' = a_1, i_1, S_1, i_m, S_2, i_1, S_3, i_m$. $\tau(S'') = \tau(S') = \tau(S)$. $\pi(S'') = \pi(a_1, i_1, T_1, i_m, T_2, i_1, T_1, i_m, T_2, i_1, T_3, i_m)$, by Q1 on i_m, $= \pi(S')$, by Q1 on i_1, $= \pi(S)$.

LEMMA 6. *If S is connected, $\alpha(S) = \alpha(S')$, $\omega(S) = \omega(S')$ and $\tau(S) = \tau(S')$, then $\pi(S) = \pi(S')$ and S' is connected.*

Proof. First note that, for any idempotents i and j, if i occurs to the right of j in S, i must be right connected to j in S' since every triple of S between this i and this j must occur in S'. But from this it follows that if i is right connected to j in S, then i is right connected to j in S'. Therefore, since S is connected, S' is connected.

Let $\alpha(S) = \alpha(S') = (a_1, i)$ and $\omega(S) = \omega(S') = j$. Then by Lemma 5, there are S_1, S_2, S_3, S_4, S_5 and S_6 such that where $T = a_1, i, S_1, j, S_2, i, S_3, j$ and $T' = a_1, i, S_4, j, S_5, i, S_6, j$, we have $\alpha(T) = \alpha(T') = \alpha(S) = \alpha(S')$, $\omega(T) = \omega(T') = \omega(S) = \omega(S')$, $\tau(T) = \tau(S_2) = \tau(S) = \tau(T') = \tau(S_5) = \tau(S')$, $\pi(T) = \pi(S)$ and $\pi(T') = \pi(S')$.

By Lemma 3, then, $\pi(a_1, i, S_4, j, S_5, i) = \pi(a_1, i, S_1, j, S_2, i)$ and $\pi(j, S_2, i, S_6, j) = \pi(j, S_2, i, S_3, j)$. Thus $\pi(T') = \pi(a_1, i, S_1, j, S_2, i, S_6, j) = \pi(T)$. So $\pi(S) = \pi(S')$.

Two distinct idempotents are *connected* in a string if each is right connected to the other. For completeness we say that, always, an idempotent is *connected* to itself (even if it appears only once in the string).

LEMMA 7. *Connectedness is an equivalence relation among the idempotents in a string.*

Proof. Reflexivity and symmetry are obvious. The transitivity of connectedness is a consequence of the transitivity of right connectedness, which is a direct consequence of the definition.

If the equivalence classes of the idempotents of a string under the connectedness relation are $I_1, I_2, \cdots, I_p$ then, for any distinct I_x and I_y, if any occurrence of an idempotent of I_x is to the left of any occurrence of an idempotent of I_y then every occurrence of every idempotent of I_x is to the left of every occurrence of every idempotent of I_y: we can then say that I_x is *to the left of* I_y.

We are now ready to complete the proof of Theorem 4.1. Suppose $\alpha(S) = \alpha(S')$, $\omega(S) = \omega(S')$ and $\tau(S) = \tau(S')$. Let $I_1, \cdots, I_p$ be in left-to-right order, the equivalence classes of idempotents of S and $I'_1, \cdots, I'_q$, in order, the equivalence classes of S'. Let i_{xL} and i_{xR}, respectively, be those idempotents of I_x that occur leftmost and rightmost in S. Similarly, define i'_{xL} and i'_{xR} in S'.

We observe that the connectedness relation in a string depends completely on the set of triples. (Indeed, if $i \neq j$ then i is right connected to j if and only if

there exists a sequence of triples $(k_0, b_1, k_1), (k_1, b_2, k_2), \cdots, (k_{r-1}, b_r, k_r)$ such that $i = k_r$ and $j = k_0$.) From this we conclude that $p = q$ and for each x, $I_x = I'_x$.

Now $i_{1L} = i'_{1L}$ and $i_{pR} = i'_{pR}$ since $\alpha(S) = \alpha(S')$ and $\omega(S) = \omega(S')$. For any x, $1 \leqq x \leqq p-1$, consider i_{xR} and $i_{(x+1)L}$. There must be a b_x such that $(i_{xR}, b_x, i_{(x+1)L})$ is a triple of S, and the only triple having both an idempotent of I_x and an idempotent of I_{x+1}. Call this the *link triple* in S of I_x and I_{x+1}. Then $(i_{xR}, b_x, i_{(x+1)L})$ must be a triple of S' and the only triple of S' with idempotents from I_x and I_{x+1}. Thus, for $1 \leqq x \leqq p$, $i_{xL} = i'_{xL}$ and $i_{xR} = i'_{xR}$. Furthermore, the link triples are all equal.

For $1 \leqq x \leqq p$, let S_x be the subpseudostring of S beginning with the left-most occurrence of i_{xL} and ending with the right-most occurrence of i_{xR}. Similarly, define S'_x from S, for $1 \leqq x \leqq p$.

Now by Lemma 6, for $1 \leqq x \leqq p$, $\pi(S_x) = \pi(S'_x)$. But then, where a_1 is the first term in both S and S', $\pi(S) = a_1\pi(S_1)b_1\pi(S_2)\cdots b_{p-1}\pi(S_p) = \pi(S')$. This concludes the proof of Theorem 4.1.

5. Local Testability in the Strict Sense. We introduce another operator. For $|W| \geqq k$, $\gamma_k(W)$ is the set of all length-k segments of W, left-end, right-end and interior. Thus $\gamma_k(W) = \beta_k(W) \cup \{\alpha_k(W), \omega_k(W)\}$.

THEOREM 5.1. *An event η is LTSS if and only if there exists a positive integer k and subsets α, γ, and ω of Σ^k such that, for all words W where $|W| \geqq k$: $W \notin \eta$ if and only if $\alpha_k(W) \in \alpha$ or $\omega_k(W) \in \omega$ or $\gamma_k(W) \cap \gamma \neq \varnothing$.*

In general the value of k needed to satisfy the condition of Theorem 5.1 is two more than the test length in the definition of "LTSS" in Section 1. The proof of Theorem 5.1 is straightforward once it is observed that a segment of length k-2 occurs in the interior of a word if and only if one of a certain set of $|\Sigma|^2$ segments of length k occurs some place in the word. (Thus, for example, if $\Sigma = \{a, b\}$, then $abb \in \beta_3(W)$ if and only if *aabba*, *aabbb*, *babba* or *babbb* $\in \gamma_5(W)$.) This completes the proof.

In this section, we shall assume that η is a regular event over Σ and ψ is the natural homomorphism of $\Sigma\Sigma^*$ to $S = SS(\eta)$.

We note that the family of LTSS events is not completely characterized by their syntactic semigroups; for the family is not closed under complementation, whereas η and its complement $\Sigma\Sigma^* - \eta$ have the same syntactic semigroup. For this reason, the algebraic condition in Theorem 5.2 below involves things other than the properties of S. It will turn out (Theorem 5.5) that we can determine whether η is LTSS by looking at the set $\psi(\eta)$ along with the properties of S.

A *right ideal* in a semigroup T is a set R such that $RT \subseteqq R$. A *left ideal* is a set L such that $TL \subseteqq L$. (The empty set is both a left ideal and a right ideal.) There is a slight terminological misfortune here in that the right ideal has to do with the left end of a word being tested, while the left ideal has to do with the right end. This anomaly is forced on us by well established semigroup terminology.

A *zero* z in a semigroup is an element such that $xz = zx = z$, for all x in the semigroup. It is both well known and easy to verify that a zero in a semi-

group, if it exists, is unique. Also easy to verify is that if W is a word such that $\Sigma^* W \Sigma^* \cap \eta = \varnothing$, then $\psi(W)$ is a zero of S.

THEOREM 5.2. *The event η is LTSS if and only if there exist subsets R, L, D, and N of S such that for all idempotents i of S: (1) R is a right ideal of S. L is a left ideal of S. If S has a zero z such that $z \in \psi(\Sigma\Sigma^* - \eta)$, then $D = \{z\} \subseteq R \cap L$; otherwise, $D = \varnothing$. (2) If $bia \in R - D$ then $bi \in R$. (3) If $bia \in L - D$ then $ia \in L$. (4) If $bia \in D$ then either $bi \in D$ or $ia \in D$. (5) $\psi^{-1}(N)$ is finite. (6) $\psi(\eta) = S - (R \cup L \cup N)$.*

Proof. Suppose first that (1)–(6) are satisfied. Then from (6) and from the fact that $\psi(\eta) \cap \psi(\Sigma\Sigma^* - \eta) = \varnothing$, a well known elementary property of the syntactic semigroup, we get

$$\Sigma\Sigma^* - \eta = \psi^{-1}(R) \cup \psi^{-1}(L) \cup \psi^{-1}(N). \tag{6'}$$

We take

$$\begin{aligned}
\tau_1 &= \psi^{-1}(R) - (\psi^{-1}(R)\Sigma \cup \psi^{-1}(D)),\\
\tau_2 &= \psi^{-1}(L) - (\Sigma\psi^{-1}(L) \cup \psi^{-1}(D)),\\
\tau_3 &= \psi^{-1}(D) - (\Sigma\psi^{-1}(D) \cup \psi^{-1}(D)\Sigma),\\
\tau_4 &= \psi^{-1}(N).
\end{aligned}$$

We leave it to the reader to verify, using (1), that

$$\psi^{-1}(D) = \Sigma^*\tau_3\Sigma^*,\ \psi^{-1}(R) = \tau_1\Sigma^* \cup \Sigma^*\tau_3\Sigma^*,\ \psi^{-1}(L) = \Sigma^*\tau_2 \cup \Sigma^*\tau_3\Sigma^*.$$

From these and (6′) we get

$$\Sigma\Sigma^* - \eta = \tau_1\Sigma^* \cup \Sigma^*\tau_2 \cup \Sigma^*\tau_3\Sigma^* \cup \tau_4. \tag{7}$$

We next note that if $W \in \tau_1$ then no proper prefix of W is in $\psi^{-1}(R)$. For suppose $W = X_1X_2a$, $a \in \Sigma$ and $X_1 \in \psi^{-1}(R)$. Since R is a right ideal, $X_1X_2 \in \psi^{-1}(R)$ and $W \in \psi^{-1}(R)\Sigma$, contradicting the definition of τ_1.

By a symmetric argument, if $W \in \tau_2$ then no proper suffix of W is in $\psi^{-1}(L)$. We also prove that if $W \in \tau_3$ then no proper segment of W is in $\psi^{-1}(D)$. For assume $W \in \tau_3$; then $\psi(W) = z$, the zero of S, and $D = \{z\}$. If X_2 is a proper segment of W with $X_2 \in \psi^{-1}(D)$ then either $W = aX_1X_2X_3$ or $W = X_1X_2X_3a$, where $a \in \Sigma$. In either case we would have $X_1X_2X_3 \in \psi^{-1}(D)$, since z is a zero, contradicting the definition of τ_3.

We next prove that τ_1, τ_2, τ_3, and τ_4 are finite. Suppose that τ_1 is infinite. Then there would be arbitrarily long words in τ_1, and so by Theorem 2.1 (Ramsey) we could get a $W_1W_2W_3 \in \tau_1$ such that $W_1 \neq \lambda$, $W_2 \neq \lambda$, $W_3 \neq \lambda$ and $\psi(W_2)$ is an idempotent. Thus $W_1W_2W_3 \in \psi^{-1}(R - D)$ but $W_1W_2 \notin \psi^{-1}(R)$, since W_1W_2 is a proper prefix of $W_1W_2W_3$. Thus (2) is violated with $b = \psi(W_1)$, $i = \psi(W_2)$, and $a = \psi(W_3)$. It follows that τ_1 is finite.

Similarly, the finiteness of τ_2 follows from (3). τ_4 is finite by (5). Finally, suppose τ_3 is infinite. Then, by Theorem 2.1, there would be $W_1W_2W_3 \in \tau_3$ where $W_1 \neq \lambda$, $W_2 \neq \lambda$, $W_3 \neq \lambda$, and $\psi(W_2)$ is an idempotent. $W_1W_2 \notin \psi^{-1}(D)$ and $W_2W_3 \notin \psi^{-1}(D)$, since these are both proper segments of $W_1W_2W_3$. But this contradicts (4).

Now take k to be one more than the maximum length of words in $\tau_1 \cup \tau_2 \cup \tau_3 \cup \tau_4$. Take $\alpha = \tau_1\Sigma^* \cap \Sigma^k$, $\omega = \Sigma^*\tau_2 \cap \Sigma^k$ and $\gamma = \Sigma^*\tau_3\Sigma^* \cap \Sigma^k$. By (7), for all words W of length at least k, $W \notin \eta$ if and only if $W \in \alpha\Sigma^* \cup \Sigma^*\omega \cup \Sigma^*\gamma\Sigma^*$. It follows by Theorem 5.1 that η is LTSS.

For the proof of Theorem 6.1 in the other direction, suppose that η is LTSS. Put

$$\sigma_1 = \{W | W\Sigma^* \cap \eta = \varnothing\},$$
$$\sigma_2 = \{W | \Sigma^* W \cap \eta = \varnothing\},$$
$$\sigma_3 = \{W | \Sigma^* W \Sigma^* \cap \eta = \varnothing\},$$
$$R = \psi(\sigma_1), \qquad L = \psi(\sigma_2), \qquad D = \psi(\sigma_3),$$
$$N = S - (R \cup L \cup \psi(\eta)).$$

From the fact that σ_1, σ_2 and σ_3 are unions of congruence classes (mod η) we get

$$\psi^{-1}(R) = \sigma_1, \tag{8}$$

$$\psi^{-1}(L) = \sigma_2, \tag{9}$$

$$\psi^{-1}(D) = \sigma_3. \tag{10}$$

Let k, α, γ and ω be as guaranteed by Theorem 5.1. Assume also that each of α, γ and ω is as large as it can be: that is, if $W \notin \alpha$ and $|W| = k$ then, for some $X \in \Sigma^*$, $WX \in \eta$; and if $W \notin \omega$ and $|W| = k$ then, for some $V \in \Sigma^*$, $VW \in \eta$; and if $W \notin \gamma$ and $|W| = k$ then, for some $V, X \in \Sigma^*$, $VWX \in \eta$. Clearly, $\psi(\alpha) \subseteqq R$, $\psi(\omega) \subseteqq L$, and $\psi(\gamma) \subseteqq D$.

LEMMA 1. *If $|W| \geqq k$ and $\gamma_k(W) \cap \gamma = \varnothing$ then there exist V and Y such that $VWY \in \eta$.*

Proof. Put $W = U_1X_1 = X_2U_2$, where $|U_1| = |U_2| = k$. Then, since U_1, $U_2 \notin \gamma$, there exist V_1, Y_1, V_2, Y_2 such that $V_1U_1Y_1 \in \eta$ and $V_2U_2Y_2 \in \eta$. Hence, $\alpha_k(V_1U_1) \notin \alpha$, $\omega_k(U_2Y_2) \notin \omega$, $\gamma_k(V_1U_1) \cap \gamma = \varnothing$ and $\gamma_k(U_2Y_2) \cap \gamma = \varnothing$. It follows that $\alpha_k(V_1WY_2) \notin \alpha$, $\omega_k(V_1WY_2) \notin \omega$. From $\gamma_k(W) \cap \gamma = \varnothing$ we also get $\gamma_k(V_1WY_2) \cap \gamma = \varnothing$. Since $|V_1WY_2| \geqq k$, we get $V_1WY_2 \in \eta$.

LEMMA 2. *If $|W| \geqq k$ then $\psi(W) \in D$ if and only if W has a segment of γ.*

The proof right to left follows from the definitions of γ and D. On the other hand, if $|W| \geqq k$ and W has no segment of γ then $W \notin \sigma_3$ by Lemma 1, and therefore $\psi(W) \notin D$, by (10).

LEMMA 3. *If $|W| \geqq k$ and $\psi(W) \in R - D$ then $\alpha_k(W) \in \alpha$.*

Proof. By Lemma 2, $\gamma_k(W) \cap \gamma = \varnothing$. So by Lemma 1, for some V and Y, $VWY \in \eta$. By (8), $W \in \sigma_1$, so $WY \notin \eta$; furthermore, $|WY| \geqq k$. Since $\omega_k(WY) \notin \omega$ and $\gamma_k(WY) \cap \gamma = \varnothing$, it must be that $\alpha_k(W) = \alpha_k(WY) \in \alpha$, by Theorem 5.1.

LEMMA 4. *If $|W| \geqq k$ and $\psi(W) \in L - D$ then $\omega_k(W) \in \omega$.*

The proof is by symmetry from Lemma 3.

LEMMA 5. *If $\psi(W) \in N$ then $|W| < k$.*

Proof. $\psi(W) \notin R$ so, for some X, $WX \in \eta$. $\psi(W) \notin L$ so, for some V, $VW \in \eta$.

Suppose $|W| \geqq k$. It would then follow that $\alpha_k(W) \notin \alpha$, $\omega_k(W) \notin \omega$ and $\gamma_k(W) \cap \gamma = \varnothing$, implying that $W \in \eta$, which contradicts $\psi(W) \in N$. Hence $|W| < k$.

We can now complete the proof of the theorem. (1) follows from the definition of R, L and D. To prove (2), assume $bia \in R-D$. Then let $\psi(U) = b$, $\psi(V) = i, \psi(W) = a$, where $U, V, W \in \Sigma\Sigma^*$. $\psi(UV^kW) \in R-D$ and $|UV^kW| > k$, so $\alpha_k(UV^kW) \in \alpha$ by Lemma 3. Since $\psi(\alpha) \subseteqq R$, $\psi(\alpha_k(UV^kW)) \in R$. Since R is a right ideal and UV^k has $\alpha_k(UV^kW)$ as an initial segment, $bi = \psi(UV^k) \in R$.

Similarly, (3) follows from Lemma 4 and (4) follows from Lemma 2. (5) follows directly from Lemma 5.

Finally, (6) follows by the simple algebra of sets from the definition of N, $\psi(\eta) \subseteqq S$, and $\psi(\eta) \cap (R \cup L) = \varnothing$. This completes the proof.

Note the degenerate cases of Theorem 5.2. If $D = \varnothing$ then η is a generalized definite event. (However not every generalized definite event is LTSS.) If $D = R = \varnothing$ then η is definite. Since every definite event is LTSS, it follows that we have here another decision procedure for definiteness. However, the latter is inferior to other decision procedures for definiteness (see [6] and [8]), since reference must be made to the mapping ψ to S, which is known not to be necessary. Furthermore, testing for condition (3) of Theorem 5.2 would seem to involve us in more work than in the other decision procedures. If $D = L = \varnothing$ then the event is reverse definite, and similar remarks apply. If $D = L = R = \varnothing$ then the event is cofinite; if in addition $N = \varnothing$ the event is $\Sigma\Sigma^*$. Finally, if S is simply a nilpotent extension of D then η is finite; also if $S = D$ then $\eta = \varnothing$. However, it is easy to tell a finite or cofinite event by its state graph, which need not be reduced: it is finite (or cofinite) if and only if there is no path containing a loop from the initial to a terminal (or, respectively, nonterminal) state.

We note also that if S has a zero $z \in \psi(\eta)$ then η is LTSS if and only if η is cofinite. The proof is left to the reader.

Theorem 5.2 clearly demonstrates that being LTSS is a decidable property. The following theorems are offered in order to improve on the efficiency of the decision procedure.

THEOREM 5.3. *Let* $P = S-\psi(\eta)$, $R = \{x|x \in P, xS \subseteqq P\}$, $L = \{x|x \in P, Sx \subseteqq P\}$, $N = P-(R \cup L)$, $D = \{z\}$ *if* z *is the zero of* S *and* $z \in P$, *and* $D = \varnothing$ *otherwise. Then conditions (1)–(6) of Theorem 5.2 are satisfied if and only if* η *is LTTS.*

The proof from left to right is immediate from Theorem 5.2. The proof right to left follows from the observation that, by the nature of the syntactic semigroup, R, L, D and N are exactly as defined in the necessity part (i.e., the second part) of the proof of Theorem 5.2.

Note that R and L of Theorem 5.3 are, respectively, the greatest right and left ideals contained in the set P.

THEOREM 5.4. *Let* I *be the set of all idempotents of* S. *Then (where* N *is defined in Theorem 5.3)* $\psi^{-1}(N)$ *is finite if and only if* $SIS \cap N = \varnothing$.

The proof is immediate from Theorem 2.1.

THEOREM 5.5. *One can determine whether η is LTSS by examining only the semigroup S and its subset $\psi(\eta)$.*

Proof. Note that the sets P, R, L, N and D of Theorem 5.3 are all determined given S and $\psi(\eta)$. The truth of conditions (1) and (6) of Theorem 5.2 are then automatically true. The truth of (2), (3) and (4) can be verified by the multiplication table of S. So can (5) by Theorem 5.4.

In my opinion the best decision procedure is outlined in the proof of Theorem 5.5. In verifying (2), (3) and (4), note that it is all right to restrict the check for those b, i, and a such that $bi = b$ and $ia = a$. One need verify only that for all such b, i, and a, $ba \in R - D$ implies $b \in R$, $ba \in L - D$ implies $a \in L$, and $ba \in D$ implies either $b \in D$ or $a \in D$. If S has a zero z then the last condition is equivalent to: $b \neq z$ and $a \neq z$ imply $ba \neq z$. If S has no zero, the last condition is vacuously satisfied. All three conditions are trivially or vacuously satisfied when $i = z$, which means that z need not be considered as a value for i in checking (2), (3) and (4).

Since η is LTSS implies that η is locally testable, it might be advisable in practical circumstances to test S for the QIC condition before embarking on the above procedure.

REFERENCES

[1] J. A. BRZOZOWSKI and I. SIMON, Characterizations of locally testable events, *Discrete Mathematics* **4** (1973), 243–271.

[2] N. CHOMSKY and M. P. SCHÜTZENBERGER, The algebraic theory of context-free languages, in *Computer Programming and Formal Systems* (P. Braffort and D. Hirschberg, editors), North Holland, Amsterdam, 1963, pp. 118–161.

[3] L. S. LEVY and M. FREEMAN, Finitely generated events, unpublished paper, 1970.

[4] R. MCNAUGHTON and S. PAPERT, *Counter-free Automata*, M.I.T. Press, Cambridge, Mass., 1971.

[5] YU. T. MEDVEDEV, On the class of events representable in a finite automaton (translated from the Russian), in *Sequential Machines—Selected Papers* (E. F. Moore, ed.), Addison-Wesley, 1964, pp. 215–227.

[6] M. PERLES, M. O. RABIN and E. SHAMIR, The theory of definite automata, *Trans. IEEE* **EC-12** (1963), 233–243.

[7] F. P. RAMSEY, On a problem of formal logic, *Proc. London Math. Soc.*, **30** (Ser. 2, 1928), Part 4, 338–384. Reprinted in a paperback, *The Foundations of Mathematics and other Logical Essays*, by F. P. Ramsey, Littlefield, Adams and Co., Patterson, N. J., 1960. (This book was first printed by Routledge and Kegan Paul, London, 1931).

[8] Y. ZALCSTEIN, Locally testable languages, *J. Computer Systems Science* **6** (1972), 151–167.

[9] Y. ZALCSTEIN, Locally testable semigroups, *Semigroup Forum* **5** (1973), 216–227.

[10] Y. ZALCSTEIN, Syntactic semigroups of some classes of star-free languages, *Proc. International Symposium on Theory of Automata, Languages and Programming* (IRIA, Paris), North Holland Publishing Company, Amsterdam, 1972, pp. 135–144.

(Received 20 January 1972, and, in revised form, 28 August 1972)