INFORMATION AND CONTROL 18, 140–155 (1971)

The Source as a Tool in Automata*

ZAMIR BAVEL

Department of Computer Science, University of Kansas, Lawrence, Kansas 66044

The source of a set of states of an automaton is introduced as the set of all predecessors of members of the set. The source is shown to be a valuable tool in proofs; it is also shown to provide new insights and directions in automata theory. Relationships are explored between the source on the one hand and, on the other, subautomata, primaries, blocks, connectivity, separation, strong connectedness, retrievability, Abelian automata, homomorphisms, and other concepts.

1. INTRODUCTION

The theory of automata has long been hampered by two facts: The lack of standard notation and the scarcity of basic manipulative tools. The former often promotes confusion, which is at best an inconvenience. However, the lack of tools is much more serious; not only does it result in long and cumbersome proofs, but it damages the clarity of insight and, much too often, it conceals new concepts and relationships which might advance the theory and practice.

It is hoped that this article contributes to alleviating both difficulties by presenting the source as a tool which aids in understanding the structure of automata and provides manipulative agility. As a case in point, it was the source which enabled the authors of Ref. [5] to generalize the notion of a primary from finite to arbitrary automata, which yielded a profitable reduction of problems concerning homomorphisms of automata; it was also responsible for the discovery of the source-splitting automaton, which is the only nonempty automaton without primaries.

In Ref. [1] the successor operator δ was used profitably as a set function. (Some manipulative aids from Refs. [1] and [2] are presented in Section 2.) The source operator σ, also a set function, is a complementary notion to the

* This work was supported in part by the National Science Foundation under grant GJ-639.

successor operator (although σ is not the inverse of δ), and it is hoped that these two operators will contribute to standardizing notation in automata theory, since the convenience they offer is too great to ignore.

2. Preliminaries

An *automaton* is a triple $A = (S, \Sigma, \delta)$, where S is a set (of *states*), Σ is a non-empty set (the *input alphabet*), and $\delta : S \times \Sigma^* \to S$ is the *transition function* satisfying, $\forall s \in S$ and $\forall x, y \in \Sigma^*$, $\delta(s, xy) = \delta[\delta(s, x), y]$, where Σ^* is the free monoid over Σ (the set of all strings of finite length of members of Σ, including the empty string ϵ, with $\delta(s, \epsilon) = s$, $\forall s \in S$). The symbols A, S, Σ, and δ will be used generically, when no ambiguity arises; S_B denotes the set of states of the automaton B.

A *triple* of A is an ordered triple of the form (T, Σ, δ'), where $T \subseteq S$ and δ' is the restriction of δ to $T \times \Sigma^*$ (a triple need not be closed under transitions). A triple $B = (T, \Sigma, \delta')$ is a *subautomaton* of A, written $B \leqslant A$, if and only if $\delta'(t, x) \in T$, $\forall t \in T$, $\forall x \in \Sigma^*$. We shall use δ for δ', as no ambiguity arises. The *set of successors* of $R \subseteq S$ is $\delta(R) = \{\delta(r, x) : r \in R \text{ and } x \in \Sigma^*\}$. When $R = \{r\}$, we permit $\delta(r)$ for $\delta(\{r\})$. The *automaton generated* by $R \subseteq S$ is $\langle R\rangle = (\delta(R), \Sigma, \delta)$; R is a *generating set* of (or for) $\langle R\rangle$. When $R = \{r\}$, we permit $\langle r\rangle$ for $\langle\{r\}\rangle$. An automaton A is *singly generated* if and only if $\exists s \in S$ such that $A = \langle s\rangle$, and in that event s is a generator of $\langle s\rangle$. The *set of generators* of $\langle s\rangle$ is $\text{gen}\langle s\rangle = \{t \in S_{\langle s\rangle} : \langle t\rangle = \langle s\rangle\}$. The *empty automaton* is $\langle\varnothing\rangle = (\varnothing, \Sigma, \delta)$. B is a *proper subautomaton* of A if and only if $B \leqslant A$ and $\langle\varnothing\rangle \neq B \neq A$. An automaton is *finite* if and only if its set of states is finite. (Unless otherwise stated, A is an automaton which is not necessarily finite.)

A *primary* of a nonempty finite automaton is a maximal singly generated subautomaton. A nonempty automaton A is *strongly connected* if and only if $\forall s, t \in S$, $s \in \delta(t)$. A nonempty automaton A is *retrievable* ([4]) if and only if $\forall s \in S$, $\forall x \in \Sigma^*$, $\exists y \in \Sigma^*$ such that $\delta(s, xy) = s$.

Where $A_i = (S_i, \Sigma, \delta_i) \leqslant A$, $\forall i \in I$, for some nonempty indexing set I, $\bigcup_{i\in I} A_i = (\bigcup_{i\in I} S_i, \Sigma, \delta')$ and $\bigcap_{i\in I} A_i = (\bigcap_{i\in I} S_i, \Sigma, \delta'')$, where δ' and δ'' are the respective restrictions of δ to $(\bigcup_{i\in I} S_i) \times \Sigma^*$ and $(\bigcap_{i\in I} S_i) \times \Sigma^*$.

The following results are straightforward consequences of the preceding definitions (see Ref. [2]) and thus are presented without proofs. They are used implicitly in the following sections.

Let A be an automaton. Then $\forall R, T \subseteq S$; (i) $S_{\langle R\rangle} = \delta(R)$; (ii) $R \subseteq \delta(R)$; (iii) $R \subseteq T \Rightarrow \delta(R) \subseteq \delta(T)$; (iv) $\delta(R) = R \Leftrightarrow (R, \Sigma, \delta) \leqslant A$; (v) $\delta(R \cup T) =$

$\delta(R) \cup \delta(T)$; (vi) $\delta(R \cap T) \subseteq \delta(R) \cap \delta(T)$; (vii) $\delta(R) - \delta(T) \subseteq \delta(R - T)$; (viii) $\langle R \rangle \cup \langle T \rangle = \langle R \cup T \rangle$; (ix) $\forall B, C \leqslant A$, $B = C$ if and only if $B \leqslant C$ and $C \leqslant B$.

The union and the intersection of subautomata of A are themselves subautomata of A.

3. The Source

Intuitively, the concept of source is based on that of a predecessor. A state t is a predecessor of a state s if and only if s can be reached from t (by a finite input sequence, including the empty sequence ϵ). A state t is a predecessor of a subset R of states if and only if t is a predecessor of some member of R. The source of R is then the set of predecessors of R. A somewhat sharper tool, and a more general concept, results from restricting the predecessors to a subautomaton. Thus, where $B \leqslant A$, the B source of R is the set of predecessors of R which are states of B.

Definition 1. Let $B = (T, \Sigma, \delta) \leqslant A$ and let $R \subseteq T$. The B *source* of R is

$$\sigma_B(R) = \{t \in T : \delta(t, x) \in R, \text{ for some } x \in \Sigma^*\}.$$

We permit $\sigma_B(r)$ for $\sigma_B(\{r\})$. We also permit the use of "source" for "A source" and "$\sigma(R)$" for "$\sigma_A(R)$" when A is the parent automaton and no confusion arises.

The interchangeability of inputs in an automaton is a rather special property, but a very convenient one. Several researchers have investigated various aspects of Abelian automata; such studies may be enhanced by the added information on the structure of such automata, which is provided in Theorem 1.

An automaton $A = (S, \Sigma, \delta)$ is said to be *Abelian* if and only if $\forall s \in S$, $\forall x, y \in \Sigma^*$, $\delta(s, xy) = \delta(s, yx)$.

Theorem 1. *In an Abelian automaton, the sets of states of disjoint subautomata have disjoint sources.*

Proof. Let A be Abelian, let $B, C \leqslant A$, and let $B \cap C = \langle \varnothing \rangle$. Suppose $s \in \sigma(S_B) \cap \sigma(S_C)$. Then $\exists x, y \in \Sigma^*$ such that $\delta(s, x) \in S_B$ and $\delta(s, y) \in S_C$. Since $B, C \leqslant A$, $\delta(s, xy) \in S_B$ and $\delta(s, yx) \in S_C$. But since A is Abelian, $\delta(s, xy) = \delta(s, yx)$, contradicting the disjointness of B and C. ∎

Although Abelian automata may not be of wide-spread interest, they serve to illustrate the usefulness of the source as a tool. We return to this topic in Section 7 with an illustration of relationships of wider applicability.

The following paraphrase of Definition 1 is often more useful and is used interchangeably with the definition.

LEMMA 1. *Under the conditions of Definition* 1,

$$\sigma_B(R) = \{t \in S_B : \delta(t) \cap R \neq \varnothing\}.$$

The proof is immediate from Definition 1.

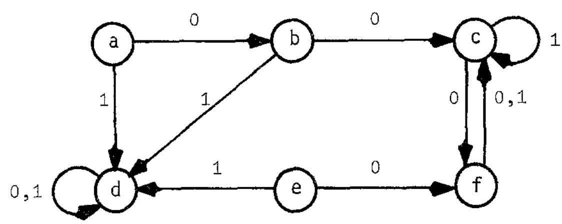

FIG. 1. State diagram of A.

As an illustration, for the automaton whose state diagram is shown in Fig. 1, $\sigma_{\langle d\rangle}(d) = \{d\}$, $\sigma_{\langle e\rangle}(d) = \{d, e\}$, $\sigma_{\langle b\rangle}(d) = \{b, d\}$, $\sigma_{\langle a\rangle}(d) = \{a, b, d\}$, $\sigma(d) = \sigma_A(d) = \{a, b, d, e\} = \sigma_{\langle\{a,e\}\rangle}(d)$, $\sigma_{\langle a\rangle}(b) = \{a, b\}$, and $\sigma(\{e, f\}) = \{a, b, c, e, f\}$.

The concept of the source of R in a subautomaton B of A, where R is not a subset of S_B, is more general but not much more useful; the present definition suffices for our purpose.

In Lemma 2 there are collected several basic facts about the source, some of which are used implicitly in the remainder of the article.

LEMMA 2. *Let $B \leqslant A$, let $\mathscr{R}$ be a collection of subsets of S_B, and let $R, T \in \mathscr{R}$. Then*

(i) $\sigma_B(\varnothing) = \varnothing$;

(ii) $\sigma_B(S_B) = S_B$;

(iii) $R \subseteq \sigma_B(R)$;

(iv) $\sigma_B(R) = \sigma_B(\sigma_B(R))$;

(v) $R \subseteq T \Rightarrow \sigma_B(R) \subseteq \sigma_B(T)$;

(vi) *if $C \leqslant B$ and $R \subseteq S_C$, then $\sigma_C(R) \subseteq \sigma_B(R)$*;

(vii) $\sigma_B(R) = \sigma(R) \cap S_B$;

(viii) $\sigma_B(\bigcup \{R \in \mathscr{R}\}) = \bigcup \{\sigma_B(R) : R \in \mathscr{R}\}$;

(ix) $\sigma_B(\bigcap \{R \in \mathscr{R}\}) \subseteq \bigcap \{\sigma_B(R) : R \in \mathscr{R}\}$;
(x) $\sigma_B(T) - \sigma_B(R) \subseteq \sigma_B(T - R)$;
(xi) $S_B - \sigma_B(R) \subseteq \sigma_B(S_B - R)$.

Proof. The first six parts are immediate from the definition. (vii): $s \in \sigma_B(R) \Leftrightarrow s \in S_B$ and $\delta(s) \cap R \neq \varnothing \Leftrightarrow s \in S_B$ and $s \in \sigma(R) \Leftrightarrow s \in S_B \cap \sigma(R)$. (viii): $s \in \sigma_B(\bigcup \{R \in \mathscr{R}\}) \Leftrightarrow s \in S_B$ and $\delta(s) \cap (\bigcup \{R \in \mathscr{R}\}) \neq \varnothing \Leftrightarrow s \in S_B$ and $\bigcup \{\delta(s) \cap R : R \in \mathscr{R}\} \neq \varnothing \Leftrightarrow s \in S_B$ and $\delta(s) \cap R \neq \varnothing$ for some $R \in \mathscr{R} \Leftrightarrow s \in \sigma_B(R)$ for some $R \in \mathscr{R} \Leftrightarrow s \in \bigcup \{\sigma_B(R) : R \in \mathscr{R}\}$. (ix): $s \in \sigma_B(\bigcap \{R \in \mathscr{R}\}) \Rightarrow s \in S_B$ and $\delta(s) \cap (\bigcap \{R \in \mathscr{R}\}) \neq \varnothing \Rightarrow s \in S_B$ and $\delta(s) \cap R \neq \varnothing$, $\forall R \in \mathscr{R} \Rightarrow s \in \sigma_B(R)$, $\forall R \in \mathscr{R} \Rightarrow s \in \bigcap \{\sigma_B(R) : R \in \mathscr{R}\}$. (x) is a known property of functions and (xi) is a special case of (x). ∎

In Lemma 1, the source σ was expressed in terms of the successor function δ. Similarly, δ may be expressed in terms of the source.

LEMMA 3. *Let $B \leqslant A$ and let $R \subseteq S_B$. Then*

$$\delta(R) = \{s \in S : \sigma_B(s) \cap R \neq \varnothing\} = \{s \in S_B : \sigma_B(s) \cap R \neq \varnothing\}.$$

Proof. $R \subseteq S_B \Rightarrow S_B \cap R = R$. Hence, by Lemma 2(vii), $\sigma_B(s) \cap R = \sigma(s) \cap S_B \cap R = \sigma(s) \cap R$. Now $s \in \delta(R) \Leftrightarrow \exists r \in R$, $\exists x \in \Sigma^*$, $\delta(r, x) = s \Leftrightarrow \exists r \in R \cap \sigma(s) \Leftrightarrow \sigma(s) \cap R \neq \varnothing$ and the first equality follows. Since B is closed under transitions, $s \notin S_B \Rightarrow \sigma_B(s) = \varnothing$, and hence the second equality follows. ∎

The ability to replace an expression involving the source by one involving the successor set, and vice versa, is often helpful. The basis for such replacements is the subject of Theorem 2 and its corollaries.

THEOREM 2. *Let $B \leqslant A$, $T \subseteq S$, and $R \subseteq S_B$. Then $\sigma_B(R) \cap T = \varnothing \Leftrightarrow R \cap \delta(S_B \cap T) = \varnothing$.*

Proof. Let $s \in \sigma_B(R) \cap T$. By Lemma 1, $s \in S_B$ and $\delta(s) \cap R \neq \varnothing$. But then $s \in S_B \cap T$, implying $\delta(s) \subseteq \delta(S_B \cap T)$, and thus $R \cap \delta(S_B \cap T) \neq \varnothing$. Conversely, let $s \in R \cap \delta(S_B \cap T)$. By Lemma 3, $\sigma_B(s) \cap T \neq \varnothing$ and hence $\sigma_B(R) \cap T \neq \varnothing$. ∎

The following two corollaries are immediate from the theorem.

COROLLARY 1. *If, in addition to the conditions of Theorem 2, $T \subseteq S_B$, then $\sigma_B(R) \cap T = \varnothing \Leftrightarrow R \cap \delta(T) = \varnothing$.*

COROLLARY 2. *Let A be an automaton and let $R, T \subseteq S$. Then $\sigma(R) \cap T = \varnothing \Leftrightarrow R \cap \delta(T) = \varnothing$.*

COROLLARY 3. *Let A be an automaton and let $R, T \subseteq S$. Then*

(i) $R \cap \delta(\sigma(T)) = \varnothing \Leftrightarrow \delta(\sigma(R)) \cap T = \varnothing$;
(ii) $R \cap \sigma(\delta(T)) = \varnothing \Leftrightarrow \sigma(\delta(R)) \cap T = \varnothing$.

Proof. Each of (i) and (ii) results from two applications (in opposite directions) of Corollary 2. ▮

Theorem 2 and its corollaries are often used implicitly in proofs in the remainder of this article.

The two sets, $\delta(\sigma(R))$ and $\sigma(\delta(R))$, mentioned in Corollary 3, are not necessarily comparable by inclusion, as the reader may easily verify. However, when R is the set of states of a subautomaton, we do have the following result.

LEMMA 4. *Let A be an automaton and let $C = (R, \Sigma, \delta) \leqslant B \leqslant A$. Then $\sigma_B(\delta(R)) \subseteq \delta(\sigma_B(R))$.*

The proof is immediate from the fact that $\delta(R) = R$. (The set $\delta(\sigma(R))$, for $R \neq \varnothing$, has special significance, which we explore in Section 5.)

We conclude this section with two characterizations of subautomata by means of the source. The closure of the set of states of a subautomaton under the transition function is most easily expressed in terms of the successor function; that is, $\forall R \subseteq S$, $(R, \Sigma, \delta) \leqslant A$ if and only if $\delta(R) = R$. It is not surprising that a characterization of subautomata in terms of the source focuses attention on the complement $S - R$ of the pertinent set R.

THEOREM 3. *Let A be an automaton and let $R \subseteq S$. Then the following are equivalent*:

(i) *the triple (R, Σ, δ) of A is a subautomaton of A*;
(ii) $\sigma(S - R) = S - R$;
(iii) $\forall s \in S, (\sigma(s) \cap R \neq \varnothing \Rightarrow s \in R)$.

Proof. (i) $\Leftrightarrow$ (ii): $\sigma(S-R) = S-R \Leftrightarrow \sigma(S-R) \subseteq S-R \Leftrightarrow \sigma(S-R) \cap R = \varnothing \Leftrightarrow (S-R) \cap \delta(R) = \varnothing \Leftrightarrow \delta(R) \subseteq R \Leftrightarrow \delta(R) = R \Leftrightarrow (R, \Sigma, \delta) \leqslant A$. (ii) $\Leftrightarrow$ (iii): $[\forall s \in S, (\sigma(s) \cap R \neq \varnothing \Rightarrow s \in R)] \Leftrightarrow [\forall s \in S, (s \in S - R \Rightarrow \sigma(s) \cap R = \varnothing)] \Leftrightarrow \sigma(S - R) \subseteq S - R \Leftrightarrow \sigma(S - R) = S - R$. ▮

4. THE PURE SOURCE

The set of states from which nonempty input sequences lead to a set R of states is often of interest, as is the set of successors of R by nonempty input sequences. These two concepts provide sharper tools for cases where the empty input sequence is to be avoided, such as the investigation of some reversibility types ([4]). Here, however, only the definitions and a few basic elementary results are stated, including characterizations of a reflexive automaton as one illustration. These two tools are quite potent, but their development largely parallels those of the successor and source operators; therefore, we are content with an intimation, rather than a full treatment.

DEFINITION 2. Let A be an automaton, $B \leqslant A$, and $R \subseteq S_B$. The *set of pure successors* of $T \subseteq S$ is $\delta^+(T) = \{\delta(t, x) : t \in T, x \in \Sigma^* - \{\epsilon\}\}$. $\delta^+(\varnothing) = \varnothing$, and we permit the use of "$\delta^+(t)$" for "$\delta^+(\{t\})$." The *pure source of R in B* is $\sigma_B^+(R) = \{s \in S_B : \delta^+(s) \cap R \neq \varnothing\}$. We permit the use of "$\sigma_B^+(r)$" for "$\sigma_B^+(\{r\})$." We also permit the use of "$\sigma^+(R)$" for "$\sigma_A^+(R)$" when A is the parent automaton and no confusion arises.

LEMMA 5. *Let A be an automaton, $B \leqslant A$, and $R, T \subseteq S_B$. Then*

(i) $\delta^+(R) = \{s \in S : \sigma^+(s) \cap R \neq \varnothing\} = \{s \in S_B : \sigma^+(s) \cap R \neq \varnothing\}$;
(ii) $R \subseteq T \Rightarrow \sigma_B^+(R) \subseteq \sigma_B^+(T)$;
(iii) $\sigma_B^+(R) = S_B \cap \sigma^+(R)$;
(iv) $\sigma_B(R) = \sigma_B^+(R) \cup R$;
(v) $\sigma_B^+(\sigma_B^+(R)) \subseteq \sigma_B^+(R)$;
(vi) $\sigma_B(\sigma_B^+(R)) = \sigma_B^+(\sigma_B(R)) = \sigma_B^+(R)$;
(vii) $\sigma_B^+(R \cup T) = \sigma_B^+(R) \cup \sigma_B^+(T)$;
(viii) $\sigma_B^+(R \cap T) \subseteq \sigma_B^+(R) \cap \sigma_B^+(T)$;
(ix) $\sigma_B^+(T) - \sigma_B^+(R) \subseteq \sigma_B^+(T - R)$;
(x) $\forall U \subseteq S, \delta^+(U) \cap R = \varnothing \Leftrightarrow U \cap \sigma_B^+(R) = \varnothing$.

The proofs of the parts of Lemma 5 are straightforward and similar to proofs of corresponding statements concerning the source; therefore, they are omitted.

LEMMA 6. *Let A be an automaton and $C = (R, \Sigma, \delta) \leqslant B \leqslant A$. Then*

(i) $R \subseteq \sigma_B^+(R)$;
(ii) $\sigma_B^+(R) = \sigma_B(R) = \sigma_B(\delta(R)) = \sigma_B(\delta^+(R))$.

Proof. (i): Since $\delta^+(s) \neq \varnothing$, $\forall s \in S$, and since $\forall r \in R = S_C$, $\delta^+(r) \subseteq \delta(r) \subseteq \delta(R) = R$, we have $\delta^+(r) \cap R \neq \varnothing$, $\forall r \in R$. Hence, $R \subseteq \sigma_B^+(R)$, by Definition 2. (ii): Since $\sigma_B^+(R) \subseteq \sigma_B(R) \subseteq \sigma_B(\delta^+(R)) \subseteq \sigma_B(\delta(R))$, it is sufficient to show that $\sigma_B(\delta(R)) \subseteq \sigma_B^+(R)$. Since $R = S_C$, $\delta(R) = R$ and hence $\sigma_B(\delta(R)) = \sigma_B(R)$. By Lemma 5(iv) and Lemma 6(i), $\sigma_B(R) = \sigma_B^+(R) \cup R = \sigma_B^+(R)$. ∎

DEFINITION 3. An automaton A is *reflexive* if and only if $\forall s \in S$, $\exists x \in \Sigma^*$ such that $x \neq \epsilon$ and $\delta(s, x) = s$.

THEOREM 4. *Let A be an automaton. Then the following are equivalent:*

(i) *A is reflexive;*
(ii) $\forall s \in S$, $s \in \sigma^+(s)$;
(iii) $\forall B \leqslant A$, $\forall R \subseteq S_B$, $R \subseteq \sigma_B{}^+(R)$;
(iv) $\forall B \leqslant A$, $\forall R \subseteq S_B$, $R \subseteq \delta^+(\sigma_B{}^+(R))$;
(v) $\forall B, C \leqslant A$, $[C \leqslant B \leqslant A \Rightarrow S_C \subseteq \delta^+(\sigma_B{}^+(S_C))]$.

Proof. (i) $\Rightarrow$ (ii): Immediate from Definition 3. (ii) $\Rightarrow$ (iii): Let $B \leqslant A$, $R \subseteq S_B$, and $s \in R$. Since $s \in \sigma^+(s)$, $s \in \sigma_B{}^+(s)$ by Lemma 5(iii), and hence, by Lemma 5(ii), $s \in \sigma_B{}^+(R)$. Thus, $R \subseteq \sigma_B{}^+(R)$. (iii) $\Rightarrow$ (iv): Let $B \leqslant A$, $R \subseteq S_B$, and $s \in R$. With $R = \{s\}$ in (iii), we have $s \in \sigma_B{}^+(s) \subseteq \sigma_B{}^+(R)$, implying $\sigma_B{}^+(s) \cap \sigma_B{}^+(R) \neq \varnothing$ and hence, by Lemma 5(x), $s \in \delta^+(\sigma_B{}^+(R))$. (iv) $\Rightarrow$ (v): Immediate. (v) $\Rightarrow$ (i): Let $s \in S$. Then by (v), $\delta(s) \subseteq \delta^+(\sigma^+_{\langle s \rangle}(\delta(s)))$ and hence $s \in \delta^+(\sigma^+_{\langle s \rangle}(\delta(s)))$. Thus $\exists t \in \sigma^+_{\langle s \rangle}(\delta(s))$ and $x \in \Sigma^*$ such that $x \neq \epsilon$ and $\delta(t, x) = s$. But then $t \in \delta(s)$, and therefore $\exists y \in \Sigma^*$ such that $\delta(s, y) = t$. Hence $\delta(s, yx) = s$ and, since $x \neq \epsilon$, A is reflexive. ∎

The proof to the following corollary to Theorem 4 follows directly from Theorem 4 (ii) and (iii) and Lemma 5 (iv) and (x).

COROLLARY. *An automaton A is reflexive if and only if $\forall R \subseteq S$, $\delta^+(R) = \delta(R)$, if and only if $\forall B \leqslant A$, $\forall R \subseteq S_B$, $\sigma^+(R) = \sigma(R)$.*

5. SOURCE AND PRIMARIES

The set $\delta(\sigma(R))$, where $\varnothing \neq R \subseteq S$, which is the subject of Lemma 4 (Section 3, above), is an interesting and useful substructure of automata, and in particular of finite automata. Intuitively, in the finite case $\sigma(R)$ consists of all the predecessors of the states of R, among which are the generators of the maximal singly generated subautomata—the primaries—which have members of R as states. Consequently, $\delta(\sigma(R))$ consists of the set of all states of these primaries, as is shown in the following.

THEOREM 5. *Let A be a finite automaton with $\varnothing \neq R \subseteq S$. Then $\langle \sigma(R) \rangle$ is the union of all primaries of A in which at least one member of R is a state.*

Proof. Let U be the union of all primaries of A in which at least one member of R is a state, and let $\langle s \rangle$ be any such primary. Then $\exists r \in R$ such that $r \in \delta(s)$. Hence, $s \in \sigma(r) \subseteq \sigma(R)$, implying $\delta(s) \subseteq \delta(\sigma(R))$ and consequently $\langle s \rangle \leqslant \langle \sigma(R) \rangle$. Hence,

$$U \leqslant \langle \sigma(R) \rangle. \tag{1}$$

Conversely, let $t \in \delta(\sigma(R))$. Then $\exists r \in R$ such that $t \in \delta(\sigma(r))$ and hence $\sigma(t) \cap \sigma(r) \neq \varnothing$. Let $p \in \sigma(t) \cap \sigma(r)$. Then $t \in \delta(p)$ and $r \in \delta(p)$. By the Primary Decomposition Theorem ([1], Theorem 1), there exists a primary $\langle q \rangle$ of A in which p is a state. But then $p \in \delta(q)$, hence $t \in \delta(q)$ and $r \in \delta(q)$, and thus t is a state of a primary $\langle q \rangle$ in which r is also a state. Consequently, t is a state of U and hence

$$\langle \sigma(R) \rangle \leqslant U. \tag{2}$$

By (1) and (2), $\langle \sigma(R) \rangle = U$. ▮

(See the illustration in Section 6 for a method for determining $\delta(\sigma(R))$ for a nonempty set R of states of a finite automaton.)

The definition of a primary of a finite automaton A as a maximal singly generated subautomaton of A was introduced in Ref. [1]. This form of the definition, however, was not easily and profitably generalizable to arbitrary automata. The generalization in Ref. [5], which yielded the desired results on decomposition and homomorphisms of arbitrary automata, was accomplished with the aid of the characterization in Theorem 6, below, to which we lead with a definition and several intermediate results. Lemma 7 is of interest apart from its use in proving Theorem 6.

LEMMA 7. *Let A be a singly generated automaton and let $s \in \text{gen}A$. Then $\sigma(s) = \text{gen}A$.*

Proof. $t \in \sigma(s) \Leftrightarrow s \in \delta(t) \Leftrightarrow t \in \text{gen}A$. ▮

DEFINITION 4. Let A be an automaton and let $R \subseteq S$. Then R is *genetic* if and only if $\sigma(R) \subseteq \delta(R)$.

In the event R of Theorem 5 is genetic, the theorem reduces to the following.

LEMMA 8. *Let A be a finite automaton and let R be a nonempty genetic subset of S. Then $\langle R \rangle$ is the union of those primaries of A in which at least one member of R is a state.*

Proof. Since R is genetic, $\delta(\sigma(R)) \subseteq \delta(\delta(R)) = \delta(R)$. But $R \subseteq \sigma(R) \Rightarrow \delta(R) \subseteq \delta(\sigma(R))$ and hence $\delta(\sigma(R)) = \delta(R)$. Thus, the lemma follows from Theorem 5. ▮

Lemma 7 indicates that gen$A$ is a genetic set for every singly generated automaton. Indeed, every nonempty finite automaton has a genetic generating set. However, not every generating set of a subautomaton is genetic. Moreover, not every singly generated subautomaton has a genetic generating set, as the reader may easily verify. The following two lemmas show that every primary of a finite automaton has a genetic generating set and that only unions of primaries display this characteristic.

LEMMA 9. *Every primary of a finite automaton has a genetic generating set.*

Proof. Let $\langle s \rangle$ be a primary of the finite automaton A. Then by Lemma 7, $\sigma_{\langle s \rangle}(s) = \text{gen}\langle s \rangle$. But $\sigma_{\langle s \rangle}(s) = \sigma(s)$, since $\langle s \rangle$ is maximal singly generated, and hence $\sigma(s) = \text{gen}\langle s \rangle$. Now $\text{gen}\langle s \rangle \subseteq \delta(s)$, since $\delta(s) = S_{\langle s \rangle}$, and hence $\sigma(s) \subseteq \delta(s)$. Consequently, $\{s\}$ is a genetic generating set of $\langle s \rangle$. ▮

LEMMA 10. *A nonempty subautomaton C of a finite automaton A has a genetic generating set if and only if C is the union of primaries of A.*

Proof. Let $C = \bigcup_{i=1}^{k} P_i$, where P_i is a primary of A for each $i \in \{1,..., k\}$. Then each P_i has a genetic generating set R_i, by Lemma 9. Now $\sigma(\bigcup_{i=1}^{k} R_i) = \bigcup_{i=1}^{k} \sigma(R_i) \subseteq \bigcup_{i=1}^{k} \delta(R_i) = \delta(\bigcup_{i=1}^{k} R_i)$ and hence $\bigcup_{i=1}^{k} R_i$ is genetic. Also, $S_C = \bigcup_{i=1}^{k} S_{P_i} = \bigcup_{i=1}^{k} \delta(R_i) = \delta(\bigcup_{i=1}^{k} R_i)$ and hence $\bigcup_{i=1}^{k} R_i$ generates C. The converse follows from Lemma 8. ▮

THEOREM 6. *A nonempty subautomaton B of a finite automaton A is a primary of A if and only if B is a minimal nonempty subautomaton of A which has a genetic generating set.*

Proof. Let B be a primary of A. Then B has a genetic generating set, by Lemma 9, and no proper subautomaton of B has a genetic generating set, by Lemma 10. Hence, B is minimal with respect to this property.

Conversely, let B be a minimal nonempty subautomaton of A which has a genetic generating set. Then B is the union of primaries of A, by Lemma 10. If B contains more than one primary of A, each such primary is a proper subautomaton of B which has a genetic generating set, contradicting the minimality of B. Hence B is a primary of A. ▮

Further use of genetic sets, and in particular of the maximal genetic generating set of a subautomaton, is made in Ref. [5].

6. Source and Connectivity

Connected automata and separated subautomata are introduced in Ref. [6], from which we reproduce their definition.

Definition 5. A nonempty subautomaton $B = (T, \Sigma, \delta)$ of A is said to be *separated* if and only if $\delta(S - T) \cap T = \varnothing$. A nonempty automaton A is said to be *connected* if and only if it has no separated proper subautomata.

As the definition points out, the separatedness of a subautomaton B depends on whether B can be reached from outside. Reachability-from-outside, however, is a concept which is better addressed by the source than by the successor operator, as is indicated by the simplicity of the following statement.

Theorem 7. *Let A be an automaton and let $\langle\varnothing\rangle \neq B \ll A$. Then B is separated if and only if $\sigma(S_B) = S_B$.*

Proof. $\delta(S - S_B) \cap S_B = \varnothing \Leftrightarrow (S - S_B) \cap \sigma(S_B) = \varnothing \Leftrightarrow \sigma(S_B) \subseteq S_B \Leftrightarrow \sigma(S_B) = S_B$. ▮

Corollary 1. *Under the conditions of Theorem 7, B is separated if and only if $(S - S_B) \cap \sigma(S_B) = \varnothing$.*

The proof of the corollary is part of the proof of the theorem. Corollary 2 provides a double test for separatedness.

Corollary 2. *Let A be an automaton and let $\varnothing \neq R \subseteq S$. Then*

(i) $\delta(\sigma(R)) = \sigma(R) \Rightarrow (\sigma(R), \Sigma, \delta)$ *is a separated subautomaton of A*;

(ii) $\sigma(\delta(R)) = \delta(R) \Rightarrow \langle R\rangle$ *is separated.*

Proof. (i): $\delta(\sigma(R)) = \sigma(R) \Rightarrow \sigma(R)$ is the set of states of a subautomaton of A. By Lemma 2(iv), $\sigma(\sigma(R)) = \sigma(R)$ and hence, by Theorem 7, $(\sigma(R), \Sigma, \delta)$ is separated. (ii): Immediate from Theorem 7. ▮

In Ref. [6] a *block* of an automaton A is defined as a connected and separated nonempty subautomaton of A. A block is then shown to be a minimal separated, and a maximal connected, subautomaton. It is also shown that every state of a nonempty automaton A is in a block of A, and thus every automaton is the union of its blocks. Corollary 2 to Theorem 7 provides a convenient method for obtaining all the blocks, or the block containing particular states. (The algorithms for finding the source of a set, the successors of a set, the primaries with states in a specified set, and the blocks of a finite automaton are detailed in Ref. [3].)

In Theorem 7, the condition $\sigma(R) = R$ was used to characterize $B = (R, \Sigma, \delta) \leqslant A$ as a separated subautomaton of A. Stronger applications of the same condition serve to characterize retrievable (for the definition, see Section 2) and discrete automata. A discrete automaton is not of much intrinsic interest, since it is the union of one-state subautomata, or totally disconnected, but it is of considerable value in counterexamples.

DEFINITION 6. A nonempty automaton A is *discrete* if and only if $\forall s \in S$, $\delta(s) = \{s\}$.

THEOREM 8. *Let A be a nonempty automaton. Then*

(i) *A is retrievable if and only if $\sigma(S_B) = S_B$, $\forall B \leqslant A$;*

(ii) *A is discrete if and only if $\sigma(R) = R$, $\forall R \subseteq S$.*

Proof. (i): A nonempty automaton A is retrievable if and only if every nonempty subautomaton of A is separated ([6], Theorem 7); hence part (i) follows from Theorem 7. (ii): $\delta(s) = \{s\}, \forall s \in S \Leftrightarrow \{s\} = \sigma(s), \forall s \in S$. Since $\sigma(R) = R, \forall R \subseteq S \Leftrightarrow \sigma(s) = \{s\}, \forall s \in S$, part (ii) follows. ▌

As one would expect, the condition in Theorem 8(i) must be strengthened to characterize a strongly connected automaton.

THEOREM 9. *Let $\langle \varnothing \rangle \neq B \leqslant A$. Then B is strongly connected if and only if $\sigma_B(R) = S_B$ for every nonempty $R \subseteq S_B$.*

Proof. $\forall s, t \in S_B, s \in \delta(t) \Leftrightarrow \forall s, t \in S_B, t \in \sigma_B(s) \Leftrightarrow \forall s \in S_B, S_B \subseteq \sigma_B(s) \Leftrightarrow$ for all nonempty $R \subseteq S_B$, $S_B \subseteq \sigma_B(R)$. ▌

It is well known that every nonempty finite automaton has a strongly connected subautomaton. The importance of the role played by the strongly connected subautomata of a finite automaton is seen with the aid of the source, as is shown in the following.

THEOREM 10. *The set of states of a nonempty finite automaton A is the union of the sources of the sets of states of the strongly connected subautomata of A.*

Proof. Let A be a finite automaton and let $s \in S$. Then $\langle s \rangle$ has a strongly connected subautomaton B. Since $S_B \subseteq \delta(s)$, $\delta(s) \cap S_B \neq \varnothing \Rightarrow s \in \sigma(S_B)$. Since s is an arbitrary state of A, the theorem follows. ▌

7. The Shape of Finite Abelian Automata

In Theorem 1 we have shown that Abelian automata have the property that disjoint subautomata have disjoint sources. Clearly, an automaton need not be Abelian to possess this property. What, then, are the most general conditions on a finite automaton which possesses this property? The following theorem provides alternate means of describing this same property.

Theorem 11. *Let A be a nonempty finite automaton. Then the following are equivalent*:

(i) *the sets of states of disjoint subautomata have disjoint sources*;

(ii) *every primary of A has exactly one strongly connected subautomaton*;

(iii) *every block of A has exactly one strongly connected subautomaton*;

(iv) *the number of blocks of A equals the number of strongly connected subautomata of A.*

Proof. (i) $\Rightarrow$ (ii): Let $B_1, \ldots, B_k$ be all the (distinct) strongly connected subautomata of A. Then by Theorem 10, $S = \bigcup_{i=1}^{k} \sigma(S_{B_i})$. Since distinct strongly connected subautomata are disjoint, (i) implies that $\sigma(S_{B_i}) \cap \sigma(S_{B_j}) = \varnothing$ if $i \neq j$. But then $S_{B_i} \cap \delta(\sigma(S_{B_j})) = \varnothing$ and hence, by Theorem 5, the primaries containing B_j are disjoint from B_i. Consequently, since every primary has at least one strongly connected subautomaton and since B_i is an arbitrary strongly connected subautomaton of A, every primary of A has exactly one strongly connected subautomaton.

(ii) $\Rightarrow$ (i): Assume (ii) and let $B, C \leqslant A$ and $B \cap C = \langle \varnothing \rangle$. Suppose $\sigma(S_B) \cap \sigma(S_C) \neq \varnothing$; then $S_B \cap \delta(\sigma(S_C)) \neq \varnothing$ and hence, by Theorem 5, B has states in the primaries of A containing C. Let $s \in S_B \cap \delta(\sigma(S_C))$. Then there exists a primary $\langle t \rangle$ of A such that $s \in \delta(t)$ and $r \in \delta(t)$ for some $r \in S_C$. Now $\langle s \rangle$ and $\langle r \rangle$ each has a strongly connected subautomaton, and since both $\langle s \rangle$ and $\langle r \rangle$ are in the primary $\langle t \rangle$, so are their strongly connected subautomata in $\langle t \rangle$. However, by hypothesis, $\langle t \rangle$ has only one strongly connected subautomaton, and hence $\langle s \rangle \cap \langle r \rangle \neq \langle \varnothing \rangle$. But $\langle s \rangle \leqslant B$ and $\langle r \rangle \leqslant C$ and hence $B \cap C \neq \langle \varnothing \rangle$, contrary to hypothesis. Hence $\sigma(S_B) \cap \sigma(S_C) = \varnothing$.

(ii) $\Rightarrow$ (iii): Let B be a block of A, let C and D be strongly connected subautomata of B, and assume (ii). Since B is connected, there exists (by Theorem 5 of Ref. [6]) a finite sequence of primaries $P_1, \ldots, P_m$ of A such that $C \leqslant P_1$, $D \leqslant P_m$, and $P_i \cap P_{i+1} \neq \langle \varnothing \rangle$, $\forall i \in \{1, \ldots, m-1\}$. Thus, $P_1 \cap P_2$ has a strongly connected subautomaton, E. But since $E \leqslant P_1$ and

$C \leqslant P_1$, (ii) implies that $E = C$. Similarly (by induction), $P_i \cap P_{i+1}$, $\forall i \in \{1,..., m-1\}$, has a strongly connected subautomaton which must be C. Thus $C \leqslant P_m$ and $D \leqslant P_m$ and hence, by hypothesis, $C = D$. Consequently, each block of A has exactly one strongly connected subautomaton, since each block of A is the union of primaries of A and A is the union of its blocks (see Ref. [6]).

(iii) $\Rightarrow$ (ii): Let each block of A have exactly one strongly connected subautomaton. Since a primary of A must have at least one strongly connected subautomaton and since no primary can be in more than one block, each primary of A has exactly one strongly connected subautomaton (and all primaries of the same block share the same strongly connected subautomaton).

(iii) $\Leftrightarrow$ (iv): Since each block of any nonempty finite automaton has at least one strongly connected subautomaton, and since distinct blocks of A are disjoint, the equivalence of (iii) and (iv) follows. ▮

The last three conditions of Theorem 11 render a fairly vivid description of an automaton possessing the property that the sets of states of disjoint subautomata have disjoint sources: the sole strongly connected subautomaton of each block serves as a sink to which all of the primaries empty. It may still require verification whether a finite automaton is Abelian; however, a glance at the state diagram may be sufficient to tell of a finite automaton that it is not Abelian—if any block (or any primary) of the automaton has more than one strongly connected subautomaton.

8. Source and Homomorphisms

The application of the source to homomorphisms of automata is a rather extensive subject, whose thorough treatment requires a separate article. Here we present only several basic results as an attempt to whet the reader's appetite.

The pertinent definitions from Ref. [1] are collected in the following.

Definition 7. Let $A = (S, \Sigma, \delta)$ and $B = (T, \Sigma, \gamma)$ be automata. A *function* on A to B is a mapping of S to T (and the identity mapping on Σ^*). A function $f: A \to B$ is a *homomorphism* if and only if it preserves transitions by Σ^*, i.e., $\forall s \in S$, $\forall x \in \Sigma^*$, $f[\delta(s, x)] = \gamma[f(s), x]$. We denote the set of all homomorphisms on A to B by $H(A \to B)$. An *endomorphism* of A is a homomorphism on A to A. We denote the set of all endomorphisms of A by $E(A)$. An *isomorphism* on A onto B is a monic and epic homomorphism on A onto B. We denote the set of all isomorphisms on A onto B by $H(A \to B)$.

An *automorphism* of A is an isomorphism on A onto A. We denote the group of automorphisms of A by $G(A)$.

Homomorphisms do not quite preserve the source of a set but inject it into the source of the homomorphic image of the set. On the other hand, isomorphisms do preserve the source. This is the substance of the following.

THEOREM 12. *Let $A = (S, \Sigma, \delta)$ and $B = (T, \Sigma, \gamma)$ be automata, $R \subseteq S$, $f \in H(A \to B)$, and $g \in H(A \to B)$. Then*

(i) $f(\sigma_A(R)) \subseteq \sigma_B(f(R))$;

(ii) $g(\sigma_A(R)) = \sigma_B(g(R))$.

Proof. (i): $s \in \sigma_A(R) \Rightarrow \delta(s) \cap R \neq \varnothing \Rightarrow \varnothing \neq f(\delta(s) \cap R) \subseteq f(\delta(s)) \cap f(R) = \gamma(f(s)) \cap f(R) \Rightarrow f(s) \in \sigma_B(f(R))$. Hence $f(\sigma_A(R)) \subseteq \sigma_B(f(R))$.

(ii): $t \in g(\sigma_A(R)) \Leftrightarrow g^{-1}(t) \in \sigma_A(R) \Leftrightarrow \delta(g^{-1}(t)) \cap R \neq \varnothing \Leftrightarrow g^{-1}(\gamma(t)) \cap R \neq \varnothing \Leftrightarrow \gamma(t) \cap g(R) \neq \varnothing \Leftrightarrow t \in \sigma_B(g(R))$. ∎

The same relationship holds for the relative source—inclusion for homomorphisms and equality for isomorphisms.

COROLLARY. *Let A, B, f, and g be as in Theorem 12, let $C \leqslant A$, and let $R \subseteq S_C$. Then*

(i) $f(\sigma_C(R)) \subseteq \sigma_{f(C)}(f(R))$;

(ii) $g(\sigma_C(R)) = \sigma_{g(C)}(g(R))$.

Proof. (i): Since $\sigma_C(R) = S_C \cap \sigma_A(R)$, by Lemma 2(vii), $f(\sigma_C(R)) = f(S_C \cap \sigma_A(R)) \subseteq f(S_C) \cap f(\sigma_A(R)) \subseteq f(S_C) \cap \sigma_B(f(R)) = \sigma_{f(C)}(f(R))$. (ii): When f is replaced by g in the proof of (i), the two inclusions become equalities. ∎

When an endomorphism f of an automaton A maps a state of a strongly connected subautomaton B to a state of B, the f image of B is B itself (see, e.g., Ref. [2], 6.8.5). We should not expect f to map the source of B onto itself, unless $f \in G(A)$.

THEOREM 13. *Let A be an automaton, let $B \leqslant A$ be strongly connected, let $f \in G(A)$, and let $f(s) \in S_B$ for some $s \in S_B$. Then $f(\sigma(S_B)) = \sigma(S_B)$.*

Proof. Since $f(S_B) = S_B$, $f(\sigma(S_B)) = \sigma(f(S_B)) = \sigma(S_B)$, by Theorem 12. ∎

We conclude with three results which follow from the preceding ones in a straightforward manner, and hence we omit their proofs.

THEOREM 14. *Let A be an automaton, $f \in G(A)$, $s \in S$, $f(s) \in \delta(s)$, and let $\delta(s)$ be finite. Then $f(\langle s \rangle) = \langle s \rangle$.*

It is of interest to note that the conclusion of Theorem 14 is no longer valid if $\delta(s)$ is not finite.

THEOREM 15. *Let A and B be finite automata, let $C \leqslant A$ be strongly connected, let $f, g \in H(A \to B)$, and let $f(s) = g(s)$ for some $s \in \sigma_A(S_C)$. Then $f(t) = g(t)$, $\forall t \in S_C$.*

THEOREM 16. *Let A be a finite automaton, let $B \leqslant A$ be strongly connected, let $f \in E(A)$, and let $s \in \sigma(S_B)$ be such that $f(s) = s$. Then the restriction of f to B is the identity.*

It is of interest to note that, even if $f \in G(A)$ in Theorem 16, the restriction of f to $\sigma(S_B)$ need not be the identity.

RECEIVED: January 15, 1970; REVISED: August 19, 1970

REFERENCES

1. BAVEL, Z. (January 1968), Structure and transition-preserving functions of finite automata, *J. Assoc. Comput. Mach.* **15**, 135–158.
2. BAVEL, Z., *Introduction to the Theory of Automata and Sequential Machines*, unpublished.
3. BAVEL, Z., Algorithms in the structure and transition-preserving functions of finite automata, unpublished.
4. BAVEL, Z., AND MULLER, D. E. (April 1970), Connectivity and reversibility in automata, *J. Assoc. Comput. Mach.* **17**, 231–240.
5. BAVEL, Z., AND THOMAS, J. W. (1967), On the decomposability of monadic algebras and automata, 1967 IEEE Conf. Rec. on Switching and Automata Theory, 322–335.
6. BAVEL, Z., Connectivity and separation in automata, unpublished.