ON THE ALGEBRAIC THEORY OF AUTOMATA

M. P. SCHUTZENBERGER
Institut Blaise Pascal
Paris, France

A serious bibliography of automata theory (even algebraic) contains some few hundred titles which might explain my not attempting to record all the progress that has been accomplished; the quality of these works either published or awaiting publication is, as a matter of fact, an excuse to leave this task to others better qualified than myself. I believe, also, that this congress may accomplish as useful a function by profiting from the reunion to confront various opinions on the role of mathematical disciplines in the art of non-numeric information processing. Indeed, there is room on this wagon for the most varied topics and, also, for the same topic in the most diverse dress. However, in the names *cybernetics*, *information theory*, *automata theory* (or the *theory of algorithms*), the *theory of formal languages* (*grammars*) related to so many others in the titles of books and conferences, one may distinguish a common core unified, at least, by the researchers who study its multiple facets one after another. This core, it seems to me, is that part of mathematics which applies to T. I. n. N. and it is of this that I shall speak from now on, begging pardon if I should go beyond the prescribed frame. I know— must one say it?– the emptiness of pompous generalities and the ridicule of prophecies, but teaching and research forces one to adopt options which are not individual and which it is better to state explicitly. If that, which the schematism imposed by time of that which I shall attempt, provokes enough discussion, and if we learn from so many experts the hopes which make one way preferable to another, this exposé, as I would wish it, will not have been just an interlude between the conferences of high technical value enriching this congress.

Like all applications of mathematics, the theory being considered has tasks which may be regrouped as follows: to orient research by classifying the problems, by extracting the proper concepts and by unifying the arguments; to put to use the essential results accumulated by the relevant branches of mathematics; and to allow the latter to profit from a restated problematics and from intuition born of experience and of the thorough study of special cases it requires.

I doubt that there would be any disagreement on these banalities, though one may imagine many shadings in the optimism implied by the first remarks and many degrees of fervor in the pursuit of the last one. To illustrate, I shall take the theory of Krohn and Rhodes on the simulation of a finite automaton given by a cascade of elementary organs. The authors have established that the basic concept is that of the variety of monoids (more exactly, of the subgroup of the finite monoid, the automaton model). This has allowed them to use deep theorems on simple groups to characterize the modular elements necessary to such a synthesis and has led them to that which is today the best adaptation to monoids of the "set extension theorems." No one here will dispute the interest of this last result, but one will ask in what way the notion of variety of subgroups is a "good concept" and in what way this would not hold also for more tangible notions; one may even state that the theory of Krohn and Rhodes is "a fine algebraic result devoid of practical significance," since it gives no explicit detail on this last parameter. I think there is one simple answer to these two objections: the notion of minimal number of states is not a good concept because no one has yet been able to say something non-trivial about it; the notion of variety has a practical importance because it allows the formulation of non-trivial relations between describable objects in the most naive vocabulary. What is more, it is the algebraic concept which, it seems, leads in vitality and the little we know of this minimal number of states derives in a natural way from the consideration of certain varieties. Two examples: the remarkable study by Elgot and Rutledge on the minimization of incompletely specified automata is based implicitly on the discussion of *Abelian* subgroups; the theory of Trachtenbrot and McNaughton on regular expressions with no Kleene stars, blends itself with that of finite monoids whose subgroups are degenerate. The detour which has led us away from the obvious has not only revealed the underlying unity but it has, I believe, allowed a collection of substantial new results, illustrated by the work of Verbeek, Beatty or Pappert. One may ask then, if it is not rather intuitions of this sort which should be awakened in young researchers and if, without abandoning more concrete goals, one should not recognize the deplorable existence of apparently simple and ostensibly useful problems which will not submit to a head-on attack with home-made arms. True, a certain vertigo of abstraction seizes all too often those among us who prefer mathematics to its applications (the others also, as a matter of fact). Why cite those articles in which the reader must ascend pyramids of *n*-tuples to attain a few dozen "theorems" each of which could be verified in two lines if one didn't have to translate and retranslate a pile of definitions? It seems to me that there is, however, a criterion to test the validity of a problem from the point of view which concerns us here, namely, its application: it is the one I made use of above: to show non-trivial relations between objects described in a reasonably simple manner by a change of notation. For the rest, that is, the interest claimed to be practical by some or by many, let me recall, according to Moore, the history of chemistry and ask in turn whether Priestly's experiences or Lavoisier's would have attracted the attention of alchemists working on the *GREAT WORK?*

I now come to that common mathematical core which has been successfully applied to T. I. n. N. Its negative characteristics keep us from the richest provinces of analysis and arithmetic. The only instance where classical methods could have been used is that of correcting codes: the existence of module structures shown by Slepian has allowed Hocquenghem, Bose and Chaudhury to establish through galois fields the theory we know. One must then go toward fresher territory, algebra after Glushkov and his school, mathematical logic with Myhill, Wang, Medveev, Kalmar and Rabin to find ways in which to study that which seems to us as the main problems:

1. Establishing a hierarchy for questions of information processing in terms of permissible modular elements and their rules of usage.
2. The optimization in terms of the material, time or feasibility.

There is, therefore, no break: the first topic was that of classical geometry. As for the second, we should like it to encompass numerical problems and the intermediate mode of Boolean algebra studied extensively by Kuntzmann and the Grenoble group that, like circuit theory, I shall not embark on here for lack of space; after all, the considerable work of Ardeen, Rabin and Winograd leave one with the hope that selective natural hypotheses will be found to avoid the difficulties revealed by Shannon in the general instance. Therefore, if a detailed analysis is conducted of this or that problem of reliability by McCulloch or Cowan, or of sorting by Schensted, Picard or Nelson, of unusual dynamics by Holland or Arbib, by Eastman or by Neuman or A. A. Markov according to Sardinas and Patterson, subtle Gödelization by Minsky, Böhm or Manuel Blum, of complexity measure by Hartmanis, Stearns or Eggan, it is certainly for the merit of the question itself but also, as Rabin has shown, in the hope of creating from a sample case these general principles which one suspects from the positive characteristics which our common mathematical core contains:—for want of topology, the hypothesis given by usage is the finiteness of a generative or referential system. This is sufficient to escape the dull formalism of a too universal algebra and does not exclude (on the contrary, I would say, at the risk of seeming old-fashioned) that as for Buchi's fine work, concrete reality be illuminated by denying what seemed its most essential trait.

For want of geometry, what machines universally propose is the finite sequence structure of discrete elements; therefore, still, N, then free monoids. Moreover, N is a Procrustean bed to which graphs and tables must conform before being manipulated. This is the interest in setting up correlations intelligently between the most varied structures and the words and operations that combine them (see Foata's original algebras with probabilistic resonances and the generating functions of Sherman, Raney, Gross and Harrison which have us revisiting, as algebraists, chapters of classical analysis and leave us hoping to extend its methods in the manner of Magnus Fox or Lyndon to commutative cases).

That is stating also the value of these procedures for cutting and retranscription of words developed by Nivat to build a theory of compilers completing the work of Bauer and Samelson.

Finally, if I knew how to do it, I should state a third character, very manifestly linked to the non-negativeness and which sustains also other close areas of application of mathematics. Maybe there shall be a theory compensating for our present inability in spite of the works of Nerode and of Gill to act like everyone and to use widely the apparatus of linear methods and vector spaces. It is, however, questions of formal languages (another name for the parts of a free monoid) which occupy the greatest number of people and which, owing to the attention of Backus, Naur or Vauqois, Hayes or Ravzin, Markus or Benzecri, confer on our small domain a trust we should not like to fail. One must first establish the equivalences of the definitions of a family of formal languages. Mathematical logic has provided the fundamental concept of recursive insolvability which, brandished vigorously by Bar-Hillel, Ginsburg and their groups, mark for all time the boundary of certain algorithmic dreams. Even though the theorems of Markov or Lecerf have a more classical twist, no one maintains that today algebra would be anything better than a convenient formalism for families of formal languages or algebras as general as those of Yamada, Ritchie, Shepherdson and Sturgis or Kuroda; the same reason dispenses us from speaking of automatic demonstrations. It is impossible to set a precise limit but it seems not to be so for languages and systems such as those studied by Shamir, Fisher, Stahl, Hennie, Cole or Evey. At this level, the typical problem is to build an algorithm (the automaton) capable of recognizing by sequential examination of its letters if a word belongs to the given formal language; owing to Gorn, Floyd, Burks and Wright, we know how to treat similarly many other problems of ambiguity and transformation. Besides, and this is the main fact, it is generally possible to compress effectively the information accumulated during the reading of the word, therefore to identify the *states* of the automaton to classes of a regular equivalence. As Rabin, Scott and Shepherdson were first to show, the problem is only one of representative monoids and one knows the advantage that authors like Culik, Mezei, Laemmel, Deussen, Giveon or Paz have derived from that model; more complicated cases (*non-deterministic* or *probabilistic*) first require monoids with binary relations elaborated by Riguet, then by Sain and Zaretskii.

Thus, we find ourselves rich with formal languages; a remarkable ingenuity was necessary to establish radical differences between procedures to which a superficial examination would have attributed an identical power, and, to conclude, I might refer to Rado who has demonstrated so well all the benefits which the art of programming extracts from such precise and difficult problems.

But, since I want to speak for theory and for algebra, I must submit that the chance for counter examples to remain put, like that for conjectures not to remain riddles, is a function of a parameter other than non-triviality, the contingency of which is smaller than that of our efforts: the richness of their relations with the center of mathematics. It is the special merit of the structures discovered by Kleene and of those discovered by Chomsky that, having been found at so many cross-roads, they are the object of so many theorems. If the most serious authos only see the utensil virtues of finite automata and of cell memories, I must remind you that their definition, as we now know it, could be the same one as for finite monoids and free groups and I thank you for allowing me to repeat after Ṣiger: "esse autem essentiae dicit totum quod pertinet ad entitatem eius, sive potentia, sive actus, indicatum per definitionem."