COMMUNICATION AMONGST AUTOMATA [1, 2]

HEINZ VON FOERSTER, PH.D.[3]

For those of us who are actively engaged in research concerned with systems of high complexity and who think about their implementation and their future application, it is quite obvious that today we are in the midst of an era which provides the ideal conditions for the fast evolution of the automaton with mind-like behavior. Thus, I appreciate very much the occasion to be permitted to give you, the psychiatrist, a short report about these developments, because—I believe—in not too distant a future it will be the psychiatrists who will be confronted with problems arising from the interactions of man with his new baby, the "intelligent" automaton.

Furthermore, I hope that the discussion of these new machines and their potentialities will give me an appropriate vehicle to present some of the fundamental concepts in an interaction process commonly referred to as "communication," concepts which, I believe, will hold for any communication process, whether it takes place between machines, beasts and man, or between all their possible combinations.

My first task in this presentation is a rehabilitation of the term "automaton." Unfortunately, in informal discussions, but also in recent literature, journal articles and in the press, you will find the terms "automata" and "robots" freely interchanged as if they would refer to one and the same thing. This, however, is not the case. While "robot" is derived from the Czech word robotnik= worker, became popular through Capek's delightful play *Rossum's Universal Robots*, and refers to a stupid mechanism carrying out without its own initiative all that it is commanded to do, "automaton" is derived from the Greek automatizein=to act according to one's own will, and thus refers to a gadget on a much higher level of sophistication. Indeed, if you care to look up "automaton" in a dictionary(1) you will find that an automaton is ". . . a contrivance constructed to act *as if* spontaneously, through concealed motive power." It may be argued that this definition describes still a pedestrian gadget, because with patience and skill we may "reveal" the concealed mechanism. However, the situation changes drastically, if—for some reason or another—we are in principle unable to reveal that hidden mechanism. Under those circumstances we are forced to drop the "as if" in the above definition and we have a truly "free" system before us which acts on its "own will." It may, perhaps, amuse you to note that Aristotle used the term "automaton" in the latter sense(2). I presume that a bad translation in the early nineteenth century of the famous passage in "De Motu" shifted its meaning to its weaker usage of today.

At this point you may rightly ask how such systems can ever be built. Unfortunately, a tight proof of my assertion of the feasibility of such systems would take up a one-semester seminar, thus, I hope you will believe me if I can assure you that such systems can be built, if they are made up of elementary components which fulfill the following four conditions(3):

1. The elementary component must be energetically autonomous. That is, it should not receive its energy via the information channels from other such units, but should be able to extract energy from its environment.

2. The rules by which the elementary component handles the information presented to it at a particular time must depend on earlier states of the component and also on the frequency of its use.

3. The elementary component must be able to make trials. That is, it should gen-

[1] Read at the 117th annual meeting of The American Psychiatric Association, Chicago, Ill., May 8-12, 1961.

[2] The work on some aspects on automata-theory presented herein is sponsored in part by NSF Grant 17414 and ONR Contract 1834(21).

[3] Dept. of Electrical Engineering, University of Illinois, Urbana, Ill.

[Reprinted from THE AMERICAN JOURNAL OF PSYCHIATRY, Vol. 118, No. 10, April, 1962]