July 4, 1956

A FRAMEWORK FOR AN ARTIFICIAL INTELLIGENCE

By Marvin Minsky

Paper I

The following is a general description of certain components of a machine which is expected to work at the solutions of problems in a reasonably intelligent manner. Its construction is not imminent -- the system described here is intended only to provide an heuristic framework in which certain general procedures can be examined in some detail. The machine is described in terms of a diagram with several "blocks." None of these blocks are described in anything approaching complete detail; on the contrary, each of them raises its own set of theoretical problems. The value of this approach, I feel, is that such a framework makes it possible to work in the direction of relatively hard solutions to specific problems. The danger is that of being trapped into making unnecessarily elementalistic distinctions in an effort to maintain the boundaries of the boxes. But this may be a smaller risk than that involved in evading separation; <u>then</u> one is liable to find that one has n boxes, n-1 of which contain a few relays or valves, and the last behaves intelligently! I claim, at the least, that no <u>one</u> of my boxes will behave intelligently.

The idea of constructing the framework, and many details concerning the "Characterizer," the "Method box," and the "Clean-up box," originated or were agreed upon in discussions between the author and John McCarthy. Some details of the "evaluator" originated in discussion with Ray Solomonoff. Many ideas came out in meetings of the Artificial Intelligence Group as a body.

For concreteness, I will often talk as though we had a particular interest in building a machine to find proofs for theorems in mathematical logic. Actually, the machine is intended to operate over much broader classes of problems, including the making of empirical perceptions and judgments. The use of the logic-proving example is dictated not only by the convenience of the clarity of the problems involved, but by the

It would see that this device is suitable mainly for "well defined" problems o q. First k

EXPLORATION SYSTEMS AND SYNTACTIC PROCESSES

By Marvin Minsky   Paper II

A "syntactic process" is a formal system (with a set of "atomic" or "kernel" sentences and a set of transformations which with the atomic sentences generate other sentences), together with a <u>process</u> which controls the sequence in which new sentences are generated. That is, a process whose past history determines, in a specified way, which transformation is to be <u>next</u> applied to which expression(s). The process itself is not here regarded as "formal;" its behavior is involved with semantic and, frequently, empirical contingencies.

In many artificial intelligence situations the syntactic process may involve in an essential way what may be called an "exploration graph." "Problem solving" seems to usually involve such a structure. Other parts of the process may involve the use of special procedures such as "interpretations" and "examples" which may use special computers or "models." The ideas of "memory," design and performance of "empirical experiments," and computation of "characters" and "abstractions" are likely to be required. And a program for self-improvement will likely be part of any important intelligence.

These various divisions might reflect actual divisions of the machine program into parts, or even separations of hardware, or they might be just a heuristic way of thinking about a relatively interpreted operation. By the use of the term "syntactic" we wish to indicate that the emphasis is on programs in whose operation expressions are used in a manner more or less suggestive of a humanoid language.

Two topics will be examined here. The first topic will be a discussion of the exploration process for problem-solving machines. The second topic is a discussion of possible ways to construct the syntax of a (meta)language so that a formal deductive system can be used, together with a semantic, and a set of empirical verification methods. We will in fact be talking about a (meta)language which in turn can talk about a logic plus a set of interpretations, plus non-logically derived properties of the models on which the interpretations are grounded.