A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I.B.M. Corporation
C.E. Shannon, Bell Telephone Laboratories

August 31, 1955

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

The following are some aspects of the artificial intelligence problem: 1

Automatic Computers

If a machine can do a job, then an automatic calculator can be programmed to simulate the machine. The speeds and memory capacities of present computers may be insufficient to simulate many of the higher functions of the human brain, but the major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have. 2.

How Can a Computer be Programmed to Use a Language

It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning and rules of conjecture. From this point of view, forming a generalization consists of admitting a new word and some rules whereby sentences containing it imply and are implied by others. This idea has never been very precisely formulated nor have examples been worked out.

3. **Neuron Nets**

How can a set of (hypothetical) neurons be arranged so as to form concepts. Considerable theoretical and experimental work has been done on this problem by Uttley, Rashevsky and his group, Farley and Clark, Pitts and McCulloch, Minsky, Rochester and Holland, and others. Partial results have been obtained but the problem needs more theoretical work.

4. **Theory of the Size of a Calculation**

If we are given a well-defined problem (one for which it is possible to test mechanically whether or not a proposed answer is a valid answer) one way of solving it is to try all possible answers in order. This method is inefficient, and to exclude it one must have some criterion for efficiency of calculation. Some consideration will show that to get a measure of the efficiency of a calculation it is necessary to have on hand a method of measuring the complexity of calculating devices which in turn can be done if one has a theory of the complexity of functions. Some partial results on this problem have been obtained by Shannon, and also by McCarthy.

5. **Self-Improvement**

Probably a truly intelligent machine will carry out activities which may best be described as self-improvement. Some schemes for doing this have been proposed and are worth further study. It seems likely that this question can be studied abstractly as well.

6. **Abstractions**

A number of types of "abstraction" can be distinctly defined and several others less distinctly. A direct attempt to classify these and to describe machine methods of forming abstractions from sensory and other data would seem worthwhile.

7. **Randomness and Creativity**

A fairly attractive and yet clearly incomplete conjecture is that the difference between creative thinking and unimaginative competent thinking lies in the injection of a some randomness. The randomness must be guided by intuition to be efficient. In other words, the educated guess or the hunch include controlled randomness in otherwise orderly thinking.

In addition to the above collectively formulated problems for study, we have asked the individuals taking part to describe what they will work on. Statements by the four originators of the project are attached.

We propose to organize the work of the group as follows.

Potential participants will be sent copies of this proposal and asked if they would like to work on the artificial intelligence problem in the group and if so what they would like to work on. The invitations will be made by the organizing committee on the basis of its estimate of the individual's potential contribution to the work of the group. The members will circulate their previous work and their ideas for the problems to be attacked during the months preceding the working period of the group.

During the meeting there will be regular research seminars and opportunity for the members to work individually and in informal small groups.

The originators of this proposal are:

1. **C. E. Shannon**, Mathematician, Bell Telephone Laboratories. Shannon developed the statistical theory of information, the application of propositional

calculus to switching circuits, and has results on the efficient synthesis of switching circuits, the design of machines that learn, cryptography, and the theory of Turing machines. He and J. McCarthy are co-editing an Annals of Mathematics Study on "The Theory of Automata" .

2. **M. L. Minsky**, Harvard Junior Fellow in Mathematics and Neurology. Minsky has built a machine for simulating learning by nerve nets and has written a Princeton PhD thesis in mathematics entitled, "Neural Nets and the Brain Model Problem" which includes results in learning theory and the theory of random neural nets.

3. **N. Rochester**, Manager of Information Research, IBM Corporation, Poughkeepsie, New York. Rochester was concerned with the development of radar for seven years and computing machinery for seven years. He and another engineer were jointly responsible for the design of the IBM Type 701 which is a large scale automatic computer in wide use today. He worked out some of the automatic programming techniques which are in wide use today and has been concerned with problems of how to get machines to do tasks which previously could be done only by people. He has also worked on simulation of nerve nets with particular emphasis on using computers to test theories in neurophysiology.

4. **J. McCarthy**, Assistant Professor of Mathematics, Dartmouth College. McCarthy has worked on a number of questions connected with the mathematical nature of the thought process including the theory of Turing machines, the speed of computers, the relation of a brain model to its environment, and the use of languages by machines. Some results of this work are included in the forthcoming "Annals Study" edited by Shannon and McCarthy. McCarthy's other work has been in the field of differential equations.

The Rockefeller Foundation is being asked to provide financial support for the project on the following basis:

1. Salaries of $1200 for each faculty level participant who is not being supported by his own organization. It is expected, for example, that the participants from Bell Laboratories and IBM Corporation will be supported by these organizations while those from Dartmouth and Harvard will require foundation support.

2. Salaries of $700 for up to two graduate students.

3. Railway fare for participants coming from a distance.

4. Rent for people who are simultaneously renting elsewhere.

5. Secretarial expenses of $650, $500 for a secretary and $150 for duplicating expenses.

6. Organization expenses of $200. (Includes expense of reproducing preliminary work by participants and travel necessary for organization purposes.

7. Expenses for two or three people visiting for a short time.

<u>Estimated Expenses</u>

6 salaries of 1200	$7200
2 salaries of 700	1400
8 traveling and rent expenses averaging 300	2400
Secretarial and organizational expense	850
Additional traveling expenses	600
Contingencies	550
	$13,500

PROPOSAL FOR RESEARCH BY C. E. SHANNON

I would like to devote my research to one or both of the topics listed below. While I hope to do so, it is possible that because of personal considerations I may not be able to attend for the entire two months. I, nevertheless, intend to be there for whatever time is possible.

1. Application of information theory concepts to computing machines and brain models. A basic problem in information theory is that of transmitting information reliably over a noisy channel. An analogous problem in computing machines is that of reliable computing using unreliable elements. This problem has been studies by von Neumann for Sheffer stroke elements and by Shannon and Moore for relays; but there are still many open questions. The problem for several elements, the development of concepts similar to channel capacity, the sharper analysis of upper and lower bounds on the required redundancy, etc. are among the important issues. Another question deals with the theory of information networks where information flows in many closed loops (as contrasted with the simple one-way channel usually considered in communication theory). Questions of delay become very important in the closed loop case, and a whole new approach seems necessary. This would probably involve concepts such as partial entropies when a part of the past history of a message ensemble is known.

2. The matched environment - brain model approach to automata. In general a machine or animal can only adapt to or operate in a limited class of environments. Even the complex human brain first adapts to the simpler aspects of its environment, and gradually builds up to the more complex features. I propose to study the synthesis of brain models by the parallel development of a series of matched (theoretical) environments and corresponding brain models which adapt to them. The emphasis here is on clarifying the environmental model, and representing it as a mathematical structure. Often in discussing mechanized intelligence, we think of machines performing the most advanced human thought activities–proving theorems, writing music, or playing chess. I am proposing here to start at the simple and when the environment is neither hostile (merely indifferent) nor complex, and to work up through a series of easy stages in the direction of these advanced activities.

PROPOSAL FOR RESEARCH BY M. L. MINSKY

It is not difficult to design a machine which exhibits the following type of learning. The machine is provided with input and output channels and an internal means of providing varied output responses to inputs in such a way that the machine may be "trained" by a "trial and error" process to acquire one of a range of input-output functions. Such a machine, when placed in an appropriate environment and given a criterior of "success" or "failure" can be trained to exhibit "goal-seeking" behavior. Unless the machine is provided with, or is able to develop, a way of abstracting sensory material, it can progress through a complicated environment only through painfully slow steps, and in general will not reach a high level of behavior.

Now let the criterion of success be not merely the appearance of a desired activity pattern at the output channel of the machine, but rather the performance of a given manipulation in a given environment. Then in certain ways the motor situation appears to be a dual of the sensory situation, and progress can be reasonably fast only if the machine is equally capable of assembling an ensemble of "motor abstractions" relating its output activity to changes in the environment. Such "motor abstractions" can be valuable only if they relate to changes in the environment which can be detected by the machine as changes in the sensory situation, i.e., if they are related, through the structure of the environrnent, to the sensory abstractions that the machine is using.

I have been studying such systems for some time and feel that if a machine can be designed in which the sensory and motor abstractions, as they are formed, can be made to satisfy certain relations, a high order of behavior may result. These relations involve pairing, motor abstractions with sensory abstractions in such a way as to produce new sensory situations representing the changes in the environment that might be expected if the corresponding motor act actually took place.

The important result that would be looked for would be that the machine would tend to build up within itself an abstract model of the environment in which it is placed. If it were given a problem, it could first explore solutions within the internal abstract model of the environment and then attempt external experiments. Because of this preliminary internal study, these external experiments would appear to be rather clever, and the behavior would have to be regarded as rather "imaginative"

A very tentative proposal of how this might be done is described in my dissertation and I intend to do further work in this direction. I hope that by summer 1956 I wi11 have a model of such a machine fairly close to the stage of programming in a computer.

PROPOSAL FOR RESEARCH BY N. ROCHESTER

Originality in Machine Performance

In writing a program for an automatic calculator, one ordinarily provides the machine with a set of rules to cover each contingency which may arise and confront the machine. One expects the machine to follow this set of rules slavishly and to exhibit no originality or common sense. Furthermore one is annoyed only at himself when the machine gets confused because the rules he has provided for the machine are slightly contradictory. Finally, in writing programs for machines, one sometimes must go at problems in a very laborious manner whereas, if the machine had just a little intuition or could make reasonable guesses, the solution of the problem could be quite direct. This paper describes a conjecture as to how to make a machine behave in a somewhat more sophisticated manner in the general area suggested above. The paper discusses a problem on which I have been working sporadically for about five years and which I wish to pursue further in the Artificial Intelligence Project next summer.

The Process of Invention or Discovery

Living in the environment of our culture provides us with procedures for solving many problems. Just how these procedures work is not yet clear but I shall discuss this aspect of the problem in terms of a model suggested by Craik[1]. He suggests that mental action consists basically of constructing little engines inside the brain which can simulate and thus predict abstractions relating to environment. Thus the solution of a problem which one already understands is done as follows:

1.

2. The environment provides data from which certain abstractions are formed.

3. The abstractions together with certain internal habits or drives provide:

4. (a) A definition of a problem in terms of desired condition to be achieved in the future, a goal.

   (b) A suggested action to solve the problem.

   (c) Stimulation to arouse in the brain the engine which corresponds to this situation.

5. Then the engine operates to predict what this environmental situation and the proposed reaction will lead to.

6. If the prediction corresponds to the goal the individual proceeds to act as indicated.

The prediction will correspond to the goal if living in the environment of his culture has provided the individual with the solution to the problem. Regarding the individual as a stored program calculator, the program contains rules to cover this particular contingency.

For a more complex situation the rules might be more complicated. The rules might call for testing each of a set of possible actions to determine which provided the solution. A still more complex set of rules might provide for uncertainty about the environment, as for example in playing tic tac toe one must not only consider his next move but the various possible moves of the environment (his opponent).

Now consider a problem for which no individual in the culture has a solution and which has resisted efforts at solution. This might be a typical current unsolved scientific problem. The individual might try to solve it and find that every reasonable action led to failure. In other words the stored program contains rules for the solution of this problem but the rules are slightly wrong.

In order to solve this problem the individual will have to do something which is unreasonable or unexpected as judged by the heritage of wisdom accumulated by the culture. He could get such behavior by trying different things at random but such an approach would usually be too inefficient. There are usually too many possible courses of action of which only a tiny fraction are acceptable. The individual needs a hunch, something unexpected but not altogether reasonable. Some problems, often those which are fairly new and have not resisted much effort, need just a little randomness. Others, often those which have long resisted solution, need a really bizarre deviation from traditional methods. A problem whose solution requires originality could yield to a method of solution which involved randomness.

In terms of Craik's[1]S model, the engine which should simulate the environment at first fails to simulate correctly. Therefore, it is necessary to try various modifications of the engine until one is found that makes it do what is needed.

Instead of describing the problem in terms of an individual in his culture it could have been described in terms of the learning of an immature individual. When the individual is presented with a problem outside the scope of his experience he must surmount it in a similar manner.

So far the nearest practical approach using this method in machine solution of problems is an extension of the Monte Carlo method. In the usual problem which is appropriate for Monte Carlo there is a situation which is grossly misunderstood and which has too many possible factors and one is unable to decide which factors to ignore in working out analytical solution. So the mathematician has the machine making a few thousand random experiments. The results of these experiments provide a rough guess as to what the answer may be. The extension of the Monte Carlo Method is to use these results as a guide to determine what to neglect in order to simplify the problem enough to obtain an approximate analytical solution.

It might be asked why the method should include randomness. Why shouldn't the method be to try each possibility in the order of the probability that the present state of knowledge would predict for its success? For the scientist surrounded by the environment provided by his culture, it may be that one scientist alone would be unlikely to solve the problem in his life so the efforts of many are needed. If they use randomness they could all work at once on it without complete duplication of effort. If they used system they would require impossibly detailed communication. For the individual maturing in competition with other individuals the requirements of mixed strategy (using game theory terminology) favor randomness. For the machine, randomness will probably be needed to overcome the shortsightedness and prejudices of the programmer. While the necessity for randomness has clearly not been proven, there is much evidence in its favor.

The Machine With Randomness

In order to write a program to make an automatic calculator use originality it will not do to introduce randomness without using forsight. If, for example, one wrote a program so that once in every 10,000 steps the calculator generated a random number and executed it as an instruction the result would probably be chaos. Then after a certain amount of chaos the machine would probably try something forbidden or execute a stop instruction and the experiment would be over.

Two approaches, however, appear to be reasonable. One of these is to find how the brain manages to do this sort of thing and copy it. The other is to take some class of real problems which require originality in their solution and attempt to find a way to write a program to solve them on an automatic calculator. Either of these approaches would probably eventually succeed. However, it is not clear which would be quicker nor how many years or generations it would take. Most of my effort along these lines has so far been on the former approach because I felt that it would be best to master all relevant scientific knowledge in order to work on such a hard problem, and I already was quite aware of the current state of calculators and the art of programming them.

The control mechanism of the brain is clearly very different from the control mechanism in today's calculators. One symptom of the difference is the manner of failure. A failure of a calculator characteristically produces something quite unreasonable. An error in memory or in data transmission is as likely to be in the most significant digit as in the least. An error in control can do nearly anything. It might execute the wrong instruction or operate a wrong input-output unit. On the other hand human errors in speech are apt to result in statements which almost make sense (consider someone who is almost asleep, slightly drunk, or slightly feverish). Perhaps the mechanism of the brain is such that a slight error in reasoning introduces randomness in just the right way. Perhaps the mechanism that controls serial order in behavior[2] guides the random factor so as to improve the efficiency of imaginative processes over pure randomness.

Some work has been done on simulating neuron nets on our automatic calculator. One purpose was to see if it would be thereby possible to introduce randomness in an appropriate fashion. It seems to have turned out that there are too many unknown links between the activity of neurons and problem solving for this approach to work quite yet. The results have cast some light on the behavior of nets and neurons, but have not yielded a way to solve problems requiring originality.

An important aspect of this work has been an effort to make the machine form and manipulate concepts, abstractions, generalizations, and names. An attempt was made to test a theory[3] of how the brain does it. The first set of experiments occasioned a revision of certain details of the theory. The second set of experiments is now in progress. By next summer this work will be finished and a final report will have been written.

My program is to try next to write a program to solve problems which are members of some limited class of problems that require originality in their solution. It is too early to predict just what stage I will be in next summer, or just; how I will then define the immediate problem. However, the underlying problem which is described in this paper is what I intend to pursue. In a single sentence the problem is: how can I make a machine which will exhibit originality in its solution of problems?

REFERENCES

1. K.J.W. Craik, The Nature of Explanation, Cambridge University Press, 1943 (reprinted 1952), p. 92.

2. K.S. Lashley, "The Problem of Serial Order in Behavior", in Cerebral Mechanism in Behavior, the Hixon Symposium, edited by L.A. Jeffress, John Wiley & Sons, New York, pp. 112–146, 1951.

3. D. O. Hebb, The Organization of Behavior, John Wiley & Sons, New York, 1949

PROPOSAL FOR RESEARCH BY JOHN MCCARTHY

During next year and during the Summer Research Project on Artificial Intelligence, I propose to study the relation of language to intelligence. It seems clear that the direct application of trial and error methods to the relation between sensory data and motor activity will not lead to any very complicated behavior. Rather it is necessary for the trial and error methods to be applied at a higher level of abstraction. The human mind apparently uses language as its means of handling complicated phenomena. The trial and error processes at a higher level frequently take the form of formulating conjectures and testing them. The English language has a number of properties which every formal language described so far lacks.

1. Arguments in English supplemented by informal mathematics can be concise.

2. English is universal in the sense that it can set up any other language within English and then use that language where it is appropriate.

3. The user of English can refer to himself in it and formulate statements regarding his progress in solving the problem he is working on.

4. In addition to rules of proof, English if completely formulated would have rules of conjecture . The logical languages so far formulated have either been

instruction lists to make computers carry out calculations specified in advance or else formalization of parts of mathematics. The latter have been constructed so as:

1. to be easily described in informal mathematics,

2. to allow translation of statements from informal mathematics into the language,

3. to make it easy to argue about whether proofs of (???) No attempt has

been made to make proofs in artificial languages as short as informal proofs. It therefore seems to be desirable to attempt to construct an artificial language which a computer can be programmed to use on problems requiring conjecture and self-reference. It should correspond to English in the sense that short English statements about the given subject matter should have short correspondents in the language and so should short arguments or conjectural arguments. I hope to try to formulate a language having these properties and in addition to contain the notions of physical object, event, etc., with the hope that using this language it will be possible to program a machine to learn to play games well and do other tasks .

PEOPLE INTERESTED IN THE ARTIFICIAL INTELLIGENCE PROBLEM

The purpose of the list is to let those on it know who is interested in receiving documents on the problem. The people on the 1ist wlll receive copies of the report of the Dartmouth Summer Project on Artificial Intelligence. [1996 note: There was no report.]

The list consists of people who particlpated in or visited the Dartmouth Summer Research Project on Artificlal Intelligence, or who are known to be interested in the subject. It is being sent to the people on the 1ist and to a few others.

For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving.

A revised list will be issued soon, so that anyone else interested in getting on the list or anyone who wishes to change his address on it should write to:

John McCarthy
Department of Mathematics
Dartmouth College
Hanover, NH

1996 note: Not all of these people came to the Dartmouth conference. They were people we thought might be interested in Artificial Intelligence.

The list consists of:

Adelson, Marvin
Hughes Aircraft Company
Airport Station, Los Angeles, CA

Ashby, W. R.
Barnwood House
Gloucester, England

Backus, John
IBM Corporation
590 Madison Avenue
New York, NY

Bernstein, Alex
IBM Corporation
590 Madison Avenue
New York, NY

Bigelow, J. H.
Institute for Advanced Studies

Princeton, NJ

Elias, Peter
R. L. E., MIT
Cambridge, MA

Duda, W. L.
IBM Research Laboratory
Poughkeepsie, NY

Davies, Paul M.
1317 C. 18th Street
Los Angeles, CA.

Fano, R. M.
R. L. E., MIT
Cambridge, MA

Farley, B. G.
324 Park Avenue
Arlington, MA.

Galanter, E. H.
University of Pennsylvania
Philadelphia, PA

Gelernter, Herbert
IBM Research
Poughkeepsie, NY

Glashow, Harvey A.
1102 Olivia Street
Ann Arbor, MI.

Goertzal, Herbert
330 West 11th Street
New York, New York

Hagelbarger, D.
Bell Telephone Laboratories
Murray Hill, NJ

Miller, George A.
Memorial Hall

Harvard University
Cambridge, MA.

Harmon, Leon D.
Bell Telephone Laboratories
Murray Hill, NJ

Holland, John H.
E. R. I.
University of Michigan
Ann Arbor, MI

Holt, Anatol
7358 Rural Lane
Philadelphia, PA

Kautz, William H.
Stanford Research Institute
Menlo Park, CA

Luce, R. D.
427 West 117th Street
New York, NY

MacKay, Donald
Department of Physics
University of London
London, WC2, England

McCarthy, John
Dartmouth College
Hanover, NH

McCulloch, Warren S.
R.L.E., M.I.T.
Cambridge, MA

Melzak, Z. A.
Mathematics Department
University of Michigan
Ann Arbor, MI

Minsky, M. L.
112 Newbury Street

Boston, MA

More, Trenchard
Department of Electrical Engineering
MIT
Cambridge, MA

Nash, John
Institute for Advanced Studies
Princeton, NJ

Newell, Allen
Department of Industrial Administration
Carnegie Institute of Technology
Pittsburgh, PA

Robinson, Abraham
Department of Mathematics
University of Toronto
Toronto, Ontario, Canada

Rochester, Nathaniel
Engineering Research Laboratory
IBM Corporation
Poughkeepsie, NY

Rogers, Hartley, Jr.
Department of Mathematics
MIT
Cambridge, MA.

Rosenblith, Walter
R.L.E., M.I.T.
Cambridge, MA.

Rothstein, Jerome
21 East Bergen Place
Red Bank, NJ

Sayre, David
IBM Corporation
590 Madison Avenue
New York, NY

Schorr-Kon, J.J.
C-380 Lincoln Laboratory, MIT
Lexington, MA

Shapley, L.
Rand Corporation
1700 Main Street
Santa Monica, CA

Schutzenberger, M.P.
R.L.E., M.I.T.
Cambridge, MA

Selfridge, O. G.
Lincoln Laboratory, M.I.T.
Lexington, MA

Shannon, C. E.
R.L.E., M.I.T.
Cambridge, MA

Shapiro, Norman
Rand Corporation
1700 Main Street
Santa Monica, CA

Simon, Herbert A.
Department of Industrial Administration
Carnegie Institute of Technology
Pittsburgh, PA

Solomonoff, Raymond J.
Technical Research Group
17 Union Square West
New York, NY

Steele, J. E., Capt. USAF
Area B., Box 8698
Wright-Patterson AFB
Ohio

Webster, Frederick
62 Coolidge Avenue

Cambridge, MA

Moore, E. F.
Bell Telephone Laboratory
Murray Hill, NJ

Kemeny, John G.
Dartmouth College
Hanover, NH