MEMORANDUM
RM-4065-PR
APRIL 1964

COMPUTERS AND COMPREHENSION

M. Kochen, D. M. MacKay, M. E. Maron,
M. Scriven and L. Uhr

PREPARED FOR:
UNITED STATES AIR FORCE PROJECT RAND
MEMORANDUM
RM-4065-PR
APRIL 1964

COMPUTERS AND COMPREHENSION
M. Kochen, D. M. MacKay, M. E. Maron,
M. Scriven and L. Uhr

This research is sponsored by the United States Air Force under Project RAND—contract No. AF 49(638)-700 monitored by the Directorate of Development Planning, Deputy Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained in this Memorandum should not be interpreted as representing the official opinion or policy of the United States Air Force.

DDC AVAILABILITY NOTICE
Qualified requesters may obtain copies of this report from the Defense Documentation Center (DDC).
PREFACE

This Memorandum is the product of a four-week symposium on the subject of computers and comprehension, held at The RAND Corporation during the summer of 1963 as one of a series of symposia comprising a Working Session in Mathematical Biology. The present study is an examination of the problem of understanding language by an artifact.

The topic was chosen because of its importance in future computer research, which will be increasingly concerned with the design of an artifact capable of receiving and processing messages formulated in ordinary language. This line of research forms one aspect of the growing effort in fields, such as man-machine interactions and command and control systems, where automatic language-handling ability would be especially advantageous.

The Memorandum and its qualifying or dissenting footnotes simply represent views that the five symposium participants felt to be worth setting out, at least for discussion. In addition to those who are concerned with artificial intelligence as such, the Memorandum should be of interest to individuals concerned with the question of language comprehension and the meaning and interpretation of various types of linguistic utterances.

The five participants (and authors of this Memorandum) are:

Manfred Kochen, * Manager of Information Retrieval, IBM Research Laboratory, Yorktown Heights, New York;

*The participation of Manfred Kochen was supported by International Business Machines Corporation.
D. M. MacKay, Granada Research Professor of Communication, University of Keele, England;

M. E. Maron, Senior Research Staff, Computer Sciences Department, The RAND Corporation;

Michael Scriven, Professor of History and Philosophy of Science, University of Indiana, Bloomington, Indiana;

Leonard Uhr, Research Psychologist, Mental Health Institute, University of Michigan, Ann Arbor, Michigan.
SUMMARY

Understanding has both non-linguistic and linguistic aspects. An organism (natural or artificial) may be said to "understand" adequately some feature of its environment if its internal organizing system is set up to take adequate account of it (without, necessarily, having any capacity to express that understanding linguistically). In the linguistic context, a new set of capacities is needed, chiefly related to: (1) the development and use of internal models as such; (2) the distinction between linguistic and direct experiential input; and (3) the capacity to learn. For (2) it is important to distinguish between:

a) The ability to react to indicative utterances as symptoms of the state of affairs giving rise to them (i.e., as equivalent to the output of a sensory transducer);

b) The ability to perceive utterances as descriptions of that state of affairs (i.e., as semantic tools intended by their originator to set up a state of readiness for the state of affairs in question).

For understanding, whether linguistic or not, the information system must embody something equivalent to an internal model of what is understood. A model, in this sense, is distinguished from a mere mnemonic by virtue of a connection between the structure of the model and the logical relationships between the elements of what is represented. To understand the difference between a description and a mere symptom of a state of affairs, the information system must embody something equivalent to an
internal model of the communication process.

Four levels of demands for language-handling ability may be distinguished:

a) Routine ability to look up answers to specified questions, without any sign of comprehension of the tokens manipulated, even though the process may be organized and condensed by algorithms or mnemonics.

b) Ability to show an understanding of what the particular discourse is about (referring discourse to an internal model, plastic or otherwise), not of what discourse itself is (e.g., no operational distinction need here be embodied between being told something and observing it).

c) Ability to understand the intentional character of linguistic utterances as designed to be used by a motivated interlocutor. This (which requires some internal representation of the interlocutor as goal directed) makes possible evaluation of, and discourse about, the communicative process itself, when this is relevant.

d) Ability to learn, to model itself, and to engage in dialogue of a fully reciprocal character. The last demands the embodiment of a normative system accessible to linguistic address, as well as the internal "map" of the domain of discourse, including the interlocutor.

The more sophisticated features discussed in this Memorandum are required only of the highest of these levels.
FOREWORD

The specific topic of the symposium on which this Memorandum is based—computers and the comprehension of ordinary language—was chosen by M. E. Maron because of his feeling that the problem of the comprehension of natural language is vital in future computer research. First, an artifact capable of receiving and comprehending messages formulated in ordinary language could fill a growing need in a world where the written word (viz., books, journals, messages, reports, etc.) is inundating workers in science, business, and management. Secondly, it would seem that the ability to comprehend natural language would have to be a necessary characteristic of any artificial entity deserving to be called intelligent. In fact, it might be argued that only by interrogation over a wide range of subjects could one decide whether or not (or to what degree) an artifact was intelligent—and this intercommunication could be carried out effectively with a human interrogator only through the medium of ordinary language.

This Memorandum does not consider the logical design of the appropriate programs, dealing only with certain prior considerations on which the programming should be based. The manuscript was originally written by Michael Scriven as an attempt to synthesize the views expressed during the project, and has been subsequently modified in the light of further reflections by all participants. Footnotes contributed as an elaboration, dissent, or qualification are identified by the author's initials.
ACKNOWLEDGMENTS

The authors wish to express their appreciation to T. E. Harris who conceived the idea for the Working Session in Mathematical Biology, of which the present symposium was a part. The encouragement and support of this symposium given by him and by E. S. Quade, Paul Armer, and W. H. Ware are gratefully acknowledged.

We would like to also thank H. A. Simon and Allen Newell for the time they spent with us in discussing questions relating to computers and comprehension.
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREFACE</td>
<td>iii</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>v</td>
</tr>
<tr>
<td>FOREWORD</td>
<td>vii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ix</td>
</tr>
<tr>
<td>I.  INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. CRITERIA OF COMPREHENSION</td>
<td>2</td>
</tr>
<tr>
<td>Types of Criteria, 2</td>
<td></td>
</tr>
<tr>
<td>Dimensions of Understanding, 3</td>
<td></td>
</tr>
<tr>
<td>Linguistic Understanding, 4</td>
<td></td>
</tr>
<tr>
<td>III. RELATED CONCEPTS</td>
<td>11</td>
</tr>
<tr>
<td>Language and Natural Language, 11</td>
<td></td>
</tr>
<tr>
<td>Natural Language vs. Technical Languages, 13</td>
<td></td>
</tr>
<tr>
<td>Natural Language, 14</td>
<td></td>
</tr>
<tr>
<td>IV. MODELS</td>
<td>18</td>
</tr>
<tr>
<td>V. LEARNING AND MOTIVATION</td>
<td>21</td>
</tr>
<tr>
<td>VI. CONCLUSIONS</td>
<td>24</td>
</tr>
<tr>
<td>General, 24</td>
<td></td>
</tr>
<tr>
<td>Practical, 25</td>
<td></td>
</tr>
<tr>
<td>Postscript, 26</td>
<td></td>
</tr>
<tr>
<td>Appendix</td>
<td></td>
</tr>
<tr>
<td>A. A NOTE ON TERMINOLOGY (D. M. MacKay)</td>
<td>29</td>
</tr>
<tr>
<td>B. DIMENSIONS OF UNDERSTANDING (M. Scriven)</td>
<td>31</td>
</tr>
<tr>
<td>C. COMPREHENSION OF UTTERANCES (D. M. MacKay)</td>
<td>42</td>
</tr>
<tr>
<td>D. ON THE POSSIBILITY OF ENTITIES WHICH CAN BE SAID TO UNDERSTAND (M. Kochen and L. Uhr)</td>
<td>47</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>58</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

In recent years, attention in computer research has been shifting from high-speed scientific and business data processing to more sophisticated uses for computing machines. Computers have been programmed to play games, to derive theorems in logic, and to answer questions on the basis of information previously stored in memory. Some of the latter types of computer programs have been described in the literature as "understanding" or "comprehending."* There are obviously many differences between computers so programmed and a human being who would be described in the same terms. The purpose of this Memorandum is to clarify and examine these differences, and to decide which, if any, are fundamental to the concept of comprehending a natural language.

The Memorandum is concerned with the following: first, it considers criteria for comprehension, with discussion on comprehension in general (chiefly in Appendix B) and on linguistic comprehension in particular; secondly, it is concerned with some key concepts that enter into these criteria, viz., language, models, and learning. The Memorandum concludes with a brief discussion of general and practical implications.

* These terms are here taken to be synonymous.
II. CRITERIA OF COMPREHENSION

TYPES OF CRITERIA

In a well-known application of behaviorism to computers,* Turing suggested that cross-examination of an entity through a communication channel is the only legitimate test procedure for evaluating its mental capacity. On this view, if there are no questions the answers to which distinguish a man from a machine, there is no basis for withholding personal language in talking about it.

Other views make physical composition and structure the determining factor. The present authors feel that, although output behavior is crucial, it is entirely legitimate to examine the way in which it is produced, including the structure and its information flow-map, as a way of checking that it is "genuine" and not "rigged" or accidental.\(^{(1,2)}\)

On the other hand, it seems unreasonable to make protein composition a prerequisite. If it were possible to produce a human child (in protein) by artificial means, it would be thought implausible to argue that such a being was not really thinking for himself as he solved novel environmental problems of the school room just like his classmates. If then the behavior of such a constructed human could be matched by that of a machine with the same information-processing system, but using inorganic components, it would seem hard to deny that the artificial

---

*See Appendix A for a brief discussion of terminology.
agent could think, etc.* So, requiring protein composition seems unreasonable.

DIMENSIONS OF UNDERSTANDING

A detailed logical analysis of the concept of understanding in terms of its contrasting and related concepts is given in Appendix B. Here, only a list of the contrasts discussed is given, which should be sufficient to locate the concept more precisely and indicate the necessity for careful specification of the kind of understanding that a comprehending artifact is intended to embody.†

*Note that "thinking," etc., is something we attribute not to "brains" but to "people" or "agents." It would therefore be nonsense to claim that "machines" think, if by "machine" we meant only the artificial analog of "brain." (D.M.M.)

†The proposition that x understands y can be interpreted as referring to a process or to a state obtained as the result of this process. To explicate the latter, it may be useful to consider the domains of the two variables x and y, which are arguments of the two-place predicate x understands y. The domain of y includes:

a1) Sensory inputs, as touch, smell;
a2) Motor skills, as walking, speaking;
b1) Percepts, as a Roman I;
b2) Symbols;
c1) Known concepts, as oneness, as an atom;
c2) Hitherto unknown concepts in terms of relations between known ones;
d1) Compound concepts as facts, theorems, domains of discourse;
d2) Plans for action or reflection; people in terms of their intentions and needs;
e ) Any linguistic utterance;
f ) A language;
g ) Language;
1) Understanding vs. ignorance;
2) Understanding vs. knowing;
3) Understanding a thing vs. understanding language
4) Understanding language vs. understanding what language is;
5) Understanding a message vs. understanding a topic;
6) Adequacy, depth, and range of understanding;
7) Understanding vs. memorizing.

LINGUISTIC UNDERSTANDING

The basic requirement for fully linguistic comprehension of a typical natural language, as elaborated in Appendix B, is the ability to perceive language as instrumental,* and not merely symptomatic. This is essential, first because without it the denotative relation cannot be fully understood, and also because many of the concepts expressed in natural language which relate to ways in which language itself is used (e.g., truthfulness, lying, exaggeration, etc.), would otherwise be incomprehensible.

Thus, we could understand the sounds "the sky is blue," if emitted by a photoelectrically-triggered tape recorder, purely as a symptom of the physical condition that released the tape. They could for this purpose be replaced by a

---

the domain of x includes:

a) Names of people;
b) Generic names for people or groups of people;
c) Certain animals.

Whether artifacts can also be added to the latter list is, of course, the key question of this Memorandum. (M.K.)

* This point is argued in more detail by MacKay, Ref. 3.
buzz or any other distinctive noise. Even if the photo-
cell in question were directed to the sky through a blue
filter, so that from hearing "the sky is blue" we could
correctly infer that the sky was then blue, this inference
would not require or demonstrate any linguistic understand-
ing of the sounds. For that, we must know what it would
be for an English speaker to use the expression "the sky
is blue" to serve a semantic purpose.

Note, however, that this does not mean that an ex-
pression cannot be understood linguistically unless its
current user's purpose is known. An expression can be
understood linguistically even when it is not being used,
but only mentioned. What is essential for linguistic
rather than symptomatic understanding is the perception
of the expression as conventionally having a purpose; i.e.,
as a tool, designed to serve a social function whether at
the moment used as such or not.

Some distinguishing features of specific types of
utterances from the standpoint of "mechanized comprehen-
sion" are examined in Appendix C. Statements, commands,
requests, questions, warnings, instructions, advice,
promises, and the like can all be distinguished by the
processes of evaluation appropriate to them in the comput-
ing network of the recipient.

The Understanding of Statements

The example of statements may here suffice to indicate
the operational approach we favor. We take the meaning of
a statement to be its selective function on the range of
internal states of organization (or conditional preparedness
for adaptive action) of the recipients. This is to be distinguished from the magnitude of its effect, which will depend on the divergence between the receiver's initial state and the state selected. Its standard meaning is the selection function it would conventionally have; its intended meaning and interpreted meaning are analogously defined. (4-7) Sometimes, it is useful to think of a statement as a tool whose purpose is to adjust the recipient's internal representation or model of the world. (8-10) The difference between hearing someone say the volcano has blown up and hearing (or seeing) it blow is reflected by characteristic differences in the evaluatory filtering operations in the hearer. As soon as input is recognized as being linguistic, its originator and the environment are scanned for cues to authority, reliability, etc., and only when these are established is the message allowed to exercise its adjusting functions. (Of course, further feedback occurs if incompatibility arises.) Short-circuiting can, however, occur through learning, so that single cues (e.g., trusted friend speaking) may automatically open the gate to the recipient's model-adjusting system.

Propositions Involving Reference to Sensations

It is not easy to dismiss the logical intuition that an entity incapable of a certain kind of experience cannot really understand terms referring to it. But it is clear that it can possess and know to be relevant every stateable proposition about such experiences: the blind man knows
very very well that "Grass is green" and "The sea is blue" are normally true.*

Propositions About Agency

Similarly, it can be argued that with tagging it must be possible for a computer system to act as the brain of an (artificial) agent who knows that he so acts and hence knows, in principle, how to use agency language even when it refers to personal human experience. It must be emphasized that merely tying the output of a computer to a plotter, for example, and providing an independent feedback which informs the machine of the success of its "intended" actions is only to implement agency. For an entity to understand what agency is, it requires at least the capacity to analyze its actions in terms of such concepts as intention and achievement, and relate this vocabulary to the processes of analyzing the authority of commands and the meaning of descriptions (embodied in the commands or elsewhere) which it may receive or emit. There is no great harm in saying that a plotter, like a dog, understands certain simple commands, even if it lacks these meta-skills; but understanding a natural language certainly requires them. Someone would have to build a good deal into a dog's brain before it could handle this, and it is no more

*It would be less plausible to claim that either a blind man or a computer knows what is meant by saying that grass looks green. (D.M.M., M.K., L.U., M.E.M.)
illuminating to say that the plotter "essentially embodies" what is necessary than to say that a dog or a thermostat does.

Propositions About Others

Much of a natural language contains explicit or implicit references to other people. An artificial agent which understood this aspect of language would have to have some internal representation of other agents and know of the similarities and differences between its own status and abilities and theirs. This is to some degree necessary for it to understand the use or concept of language at all; but to understand that, it would need to have little conception of the motivations of others, knowing only that they may sometimes lie or err or be wrong for other reasons. (A good program would recognize the possibility of solipsistic doubt.) One special interest of propositions about others is that they can be handled by using the self as a model.* Seeing others as analogous to oneself provides an immediate structure for storage of information about them.†

Propositions Requiring Topic Understanding

This last subsection illustrates a general class of propositions which are fully understood only by someone

*See MacKay, Ref. 9, p. 83.
†It is perhaps for this reason that primitive man used this model in his attempts to store and relate data about the external world with his anthropomorphic theories and deities. (M.S.)
with some degree of topic understanding. Topic understanding multiplies the informational impact of a proposition and a special case of this effect occurs when partial message understanding combines with topic understanding to produce complete message understanding via an hypothesis about the speaker’s intentions. This procedure is embodied in some programs that are currently referred to as "semantic." They incorporate part of what, in a comprehending entity, would be called checking alternative interpretations of ambiguous input against what is known to be true, and operating on the assumption that the consistent alternative is the correct interpretation. This relatively crude algorithm, nonetheless, considerably improves their efficiency, and a more sophisticated program, which allowed for the occasions when truth is not intended (then selecting the negation as the correct interpretation) and the occasions when truth is irrelevant, etc., would obviously improve efficiency further for certain types of input (e.g., intelligence reports).

So-called "semantic machines" (better called "truth-testing machines") are unable to distinguish experience from statements about it and, hence, cannot be said to understand anything like a natural language. Of course, they respond to coded stimuli, but so does a telegraph clacker and a supermarket cash register; calling it a code is based on purely anthropomorphic convention. Language

* See Appendix B.

† But note that they do handle their intended universes—those where the truth is spoken. (L.U.)
is an agreed device, understood and often manufactured by both sides; the number of stones we throw at apples in trees or the keys we hit on a typewriter may signify something and are correlated with the output in a non-random way, but the trees and the typewriter could not be said to "understand a language" in this interaction. For that to occur, a quite different order of operation by the recipient is required which makes the distinction between language and significant sensory input (which naturally must itself be coded), and employs the right vocabulary about it.

To the suggestion that this just means more complexity, perhaps the best reply is that complexity and organization are all that distinguish molecules from atoms, monkeys from molecules, and men from monkeys: the right kind of complexity is what we are looking for in any program design or analysis task.
III. RELATED CONCEPTS

LANGUAGE AND NATURAL LANGUAGE

It would require extensive further discussion to uncover and make plausible the case for a minimum set of sufficient conditions for a language. But it seems clear that two important necessary conditions can be given.

Referential Function

Language is related to reality (including the numerical, etc., realities of mathematics, subjective realities such as pain, and the hypothesized realities of fiction) in a conventionally recognized way which enables it to function as a medium for communication. It may, of course, be abused or used to mislead, as well as to inform, etc. The perception of clouds to windward evokes expectation of shade via a very simple stored associative connection. The expression "clouds to windward" requires an extra decoder to get its meaning and an extra set of filters and feedbacks to handle the assessment of reliability (see Appendices B and C).

The loose coupling between language and the world which distinguishes statements from symptoms is notably absent in something that has often been called a language—the representation of information in afferent nerve fibers. This is systematic but purely symptomatic, and the relations between "sender," "user," and "referent" are considerably different. The retina cannot exercise an opinion whether to tell the brain what energy has reached it and
where; the brain does not consider the possibility of deliberate deception by the retina: neither of these are agents with goals and possible goal-conflicts and, hence, independence of action. Neither knows what the other is doing with the information it receives or even that it receives it as information. There is no common learning of the language, etc. These facts are not incidental to the concept of comprehending or using a language, but central; a microwave relay station is neither using nor comprehending the coded information it handles.

**Combinatorial Possibilities**

The use of language is often a symptom of some state of the user, but language is not just a set of symptoms (what does "and" symptomatize?; "part of," "number," etc.?). It is crucial that language is a combinatorial repertoire with unlimited possible combinations whose meanings can be inferred from a finite set of "rules" governing the components' meaning. (The so-called "rules" are learned as response sets and are only partly formalizable.) Thus, "Ouch!" is either a natural sign of pain and non-linguistic, or else, to the extent that it is a conventionalized sign of the same condition, a wholly atypical linguistic utterance, a limit case, an atom which cannot be combined into molecules. *

*Some linguists (e.g., Hockett, in a private communication) and logicians (e.g., Quine, in *Word and Object*) would argue that "Ouch!" is linguistic. It is an expression of pain, quite unlike a groan or reflexive response. It could, perhaps, be combined into compounds such as, "Ouch and Ouch again," or, "Ugh, Ouch, and Goodbye." (M.K.)
Animal Languages

Parrots utter sentences, and some even utter them only on appropriate occasions. But the parrots neither use nor understand language as such, because they do not understand the words they use. A dog obeying simple commands can more easily be said to understand them since he does respond to them in several different combinations. But the key test is whether he responds appropriately* to a new combination of old words. If he does so, he is beginning to understand grammar--i.e., the structure of language; and language without structure is hardly language at all.

The "language of the bees" is an interesting intermediate case which apparently does involve recombination, but it is still not clear that the watching workers could handle novel combinations, although their present repertoire appears to be much larger than that of any dog. At least, it must be said that their language (and that of the dolphins as so far disclosed) is extremely limited in its uses (and has no apparent capacity for misuse), although it does clearly demonstrate the crucial social advantages of language.

NATURAL LANGUAGE VS. TECHNICAL LANGUAGES

As the natural language springs from the needs of a society of semi-cooperating, semi-competing, semi-romantic,

*In any general treatment, this would include not responding to meaningless combinations, and responding in the same way at one level, and differently at another, to expressions with the same meaning but made up of different words.
semi-realistic beings, so technical languages spring from certain special needs. Typically, an increase in the number of discriminations needed and the labels for them is involved; frequently, the addition of new organizations and names for them is desirable. Hence, the greater precision* and novel concepts of the mathematical sciences and the language of games come about. Such languages may in general be handled more easily than natural languages by computers, since the resolution of ambiguity typically requires only syntactical, contextual, and semantic cues and not circumstantial ones.†

NATURAL LANGUAGE

Esperanto and Basic English are synthetic languages designed for basic communication. But they contain enough concepts to require a machine with all the sophistications discussed above. They are embryonic natural languages, serving quite a different function from technical languages. If we try to get a technical or artificial language to do the job of natural language, the environment and normal

---

*While normative prescriptions for how language users ought to respond to utterances dominate in the case of technical languages (e.g., "Integrate $x^2 dx$ from 0 to 1!") and descriptive rules for how users do respond to utterances dominate in natural languages (e.g., responses to "Integrate the schools in Atlanta!"). (M.K.)

†This may be seen as largely due to the fact that such languages were designed to be written, not spoken, and only partly due to any intrinsic precision, since that is always inexhaustible, or, conversely, is for certain purposes as great in the phrase "longer than" as in "one micron longer than." (M.S.)
use will soon "contaminate" it.

So natural language may be thought of as being natural in two ways: it deals with natural needs, conflicts, objects, etc., rather than the entities of artificially constructed fields; and, it evolves naturally rather than being invented.

For two reasons, the comprehension of a natural language appears to be a good requirement to set for a computer which is supposed to be capable of understanding.* First, we have a good idea how to test the understanding of such material, whereas in an artificial language which both we and the computer have to learn, it is highly debatable what constitutes comprehension. That is, success is readily recognizable. Second, the practical applications of a computer which understands a natural language are, sometimes at least, significantly greater than those of the alternatives. That is, success is really valuable.

Natural languages in this sense have been the subject of increasing attention in recent years, in both the philosophical and the computer fields, partly because they have two striking and valuable features, which may be termed flexibility and fertility.

---

*A third reason making ordinary language of special interest as a vehicle to study comprehension in relation to automata is that it refers to internal models of reality rather than to reality itself. Thus, the term "integrated schools" evokes in listeners internal images—no two listeners having exactly the same ones. Yet, there is enough in common to all these individual internal images to make communication possible. (M.K., L.U.)
Flexibility

The function words of English—the conjunctions, participles, copulas, prepositions—have almost no associations for any of us. They are the smooth stones on the river bed, worn free of any growths by the torrent of language which they support. The same is almost true of many high-frequency terms, such as "same," "almost," "many," "high." But words such as "river," "bed," "worn," "free," "growths," are highly associative and are instantly suggested by cues which occur in a new physical or conceptual environment. We naturally and immediately understand the use of river metaphors when talking of time; for example, time "rolls on," "rushes by," "flows smoothly." And, in the same way, we understand the bed metaphor when it is applied to the river. This capacity of a natural language to describe new situations immediately and illuminatingly is its flexibility. It should be appreciated in its mundane aspects as well as its poetic. A child is understood when he calls the referee a policeman, as is Keats when he speaks of "silver snarling trumpets."

Fertility

The other side of the flexibility coin is the tendency of connotations in natural language to suggest inferences or associated ideas, some appropriate and some inappropriate; from these, new theories and new philosophical puzzles spring. The "river of time" simile suggests that we ask how fast time flows, and the problem of whether an answer is an introspective datum, is metaphysical nonsense, or
is an axiom of physical metatheory, is still unclear. We rely on flexibility and hope for fertility whenever we introduce a new concept in science using old terms--force, energy, information, genetic code, diameter of universe.

So, natural language, apart from its fuzziness (for some purposes) has also a degree of flexibility and fertility which is essential for handling the everyday events and exchanges of normal life.*

*The contrast between natural and artificial or technical languages has another aspect of some practical interest. Natural language is an excrescent symbolic approximation to a world of experience essentially richer. We try to put experience, advice, etc., into words. Generally, the result enables us to recognize the essential features of the experience, advice, etc., we meant to convey; but equally generally, it leaves us dissatisfied in some detail. In short, in natural language it is the concepts that are in command--the words have to shuffle around until they match as best they can. In artificial language, on the other hand, terms are stripped as clean as possible of associations, and stick to their defined meaning. Here, the words are in command. The concepts (ideally) are restricted to a set in one-to-one correspondence with formulations. (D.M.M.)

All natural languages seem to have certain basic syntactic features in common which may reflect some basic principles of how internal models are made or of the constituents used to make them. Prepositions seem particularly worthy of study for what they might reveal about representations of the most basic concepts. (M.K., L.U.)
IV. MODELS

This section summarizes and expands slightly what is implicit in the earlier discussion. Models can be embodied in a computer (from a slide rule up) or a computer program. For a computer system to have a model in the sense we do, it must embody a second-order capacity* to represent the model itself, to distinguish it from and relate it to reality, to update it on the basis of linguistic and experiential input, to read off required data from it. The merits of modeling are economy and fertility/flexibility: it is suggestive of hypotheses (predictive and otherwise), and is able to accommodate new input without extensive rearrangement.†

A mnemonic is a device for improving recall by attaching data in an arbitrary way to a structure that is distinguished only by being easily recalled. A model, on the other hand (in any sense stronger than "descriptive rules"),

*It might be fruitful to distinguish between two types of programs (or algorithms) and to inquire if the distinction is fundamental. The first is one designed to pass as sophisticated a comprehension test on as large a class of materials as anyone specifies. The second type of program generates programs of the first kind, with capability of receiving or calling for external information. This type of program would run indefinitely. Perhaps, for the second type of program, given any comprehension test, there exists a time at which the computer will adequately pass it. For the first type of program, this may not be the case. (M.K.)

†See: Craik, Ref. 11; MacKay, Refs. 8-10, 12-13; Peirce, Ref. 14; Uhr, Refs. 15-16.
attaches data via a systematic transformation to a structure which is chosen, not for its ease of recall, but for (a) the fact that it is understood, and (b) the fact that a systematic transformation is available which we think will work for present and future data. Today's computers which embody a model cannot be said to "have" the model they use as a model in the above sense: they incorporate no more understanding than we have of a subject for which we have only an algorithm.

Computers may have (certainly will have in future) much larger immediate or total storage capacity than a human. This would reduce one need for models, that of economy, but it would not eliminate their desirability nor affect the fact that understanding is often enhanced if coupled with ability to predict.* Ability to calculate probabilities for future events is part of understanding of the nature of similar events in the past, and mere rote storage--however large--will not help with that. To understand is to have hypotheses about what is understood, and hypotheses organize expectations as well as memory. (17-19) The urge for simplicity in hypothesis-making is an urge for predictive as well as storage efficiency. To have grasped an idea or to understand a phenomenon (to have a theory of it) is partly to have found a model of it which has enough inductive support to make one feel that the extrapolation of its model-

*Because we count being surprised by X's behavior as a sign of not having fully understood X. (M.S.)
relationship is justified so that surprises are unlikely. *

Models can be formed to represent conditions at a space-time location different from that of the agent, very different from those now surrounding him, and manipulated verbally and conceptually for purposes of long-range preparedness. † (This feature is not necessary in models of family relationships or chess which are aspatial and partly atemporal, although more-localized hypotheses can be built on them. **) 

In a computer program to handle language with this degree of intelligence, mechanisms will be required for forming the triple association between linguistic description, direct experience, and model. They should themselves be subject to description by the user, if our own language, in which we discuss such matters, is to be understood in all its aspects. The same can be said about the procedure for amalgamating, collapsing, and relating models from different sensory modalities, informants, and fields.

* Pasteur's dictum that chance favors the prepared mind implies that it takes a certain understanding to recognize the possibility of surprises in the first place. (M.K.)

† Note that here again the agent's own organizing system can serve in large measure to supply the components of the "model" (see MacKay, Refs. 8-9).

** But relations equivalent to spatial and temporal ones are already found in the programs that handle family relations and chess. (L.U.)
V. LEARNING AND MOTIVATION

It will already be clear that the capacity to learn is closely connected with the capacity to understand (see Appendix D). One may feel that a computer which had all our knowledge and models of things as they are, and our language built into it, but no learning capacity, would be just as superficial and uncomprehending as one with a gigantic but wholly unorganized memory from which it produced the answers to any question it was asked about things as they are. One might suppose that a difference lies in the greater predictive commitment of the former, but the latter is ex hypothesi in possession of every fact and consequence of the former. Is it just the greater economy of the former that appeals to us?

Is it rather the fact that the former's information-flow system is better adapted to the world, that it incorporates better understanding in a way that must eventually show up in either faster retrieval, faster updating, or larger reserve storage? At first, this seems to be the cash value of models. But, if we look closely, we see that active learning capacity is already built into the first machine. To the extent that updating capacity exists at all, and we have constantly stressed its importance, learning of one kind is already present. To the extent that it is programmed for analysis of input information and consequent re-indexing of its own storage, it shows active learning. A computer could not possibly be said to understand input if it did not alter what it subsequently output when asked for related information. So
understanding entails one kind of learning.

The superstorage computer either reduces to an inefficient form of the comprehending computer, or fails to give correct answers, when we look into its updating routines. If it doesn't update, it will output contradictions. If it does, then it has the lower (storage) level of the other, less neatly arranged, and nothing else.

Such a computer could be programmed to make an active search of literature or to perform experiments in order to "improve its mind" (and its grammar), and this might be described as a search for truth. There would even be some point in describing it as "motivated to seek truth" if its self-perception was appropriately integrated with this claim. In general, the motivation structure of a supercomputer, like its memory, may well be significantly different from that of a man; but, given the usual language of motives, it should naturally extract from any new task-statement the goal involved and relate this to any subordinate goals it may have.*

So, the comprehending computer is automatically a learning computer, to some degree.\(^{(20,21)}\) To be as intelligent as it could be--to learn as much as it could from the mistakes and successes of itself and others--would require further design improvements. But the question of whether to try to match a baby's capacities or an adult's

\*At this point, we may begin to consider the possibility of its having unconscious motivation; i.e., goal-seeking patterns in its actions which it has not yet perceived itself. (L.U.)
as a starting point for a comprehending computer is a hard one. The longer the learning process, the more likely that the final performance will continue to match ours, but, in terms of originality, this may be disadvantageous.*

*See also: Kochen, Ref. 22; MacKay, Refs. 12-13.
VI. CONCLUSIONS

GENERAL

An examination of the difference between human comprehension and current computer performance has been undertaken. The major conclusion is that there appears to be no barrier in principle to the construction of an entity (an "artificial agent") which could meet all behavioral tests for comprehension, but that this will require the integration of a wholly distinct set of capacities with those embodied in current programs.

These capacities chiefly relate to: a) the development and use of internal models as such; b) the distinction between linguistic and direct experiential input; and c) learning (these capacities are not wholly distinct). The second necessitates the existence of a processing system for linguistic input which can deal not only with the mere possibility of aberrant symptoms (which a sensory channel must be able to handle), but also with the special kinds of unreliability peculiar to language; e.g., lying, misstating, exceeding authority, misunderstanding, exaggerating, acting. In short, the would-be comprehender of a rich language cannot handle those substantial areas where these sources of error are important, nor be said to understand, in toto, what a language is, without the implicit recognition that it originates in an entity who

*It seems likely that such an entity will have to grow through its experiences. It is too complex to be designed.
uses it for certain ends. The language-comprehender must have a model of the language-producer as such.*

**PRACTICAL**

It must be said immediately that something as sophisticated as a full comprehender is unnecessary and even possibly inconvenient for many practical information-processing purposes. We can distinguish several types of artifact in terms of different demands made on them:

1) The Electronic Dictionary/Grammar/Encyclopedia (EDGE). Efficient design can provide fast look-up in large volumes of natural language text and printout of answers to questions of stereotyped form. The dictionary aspect could be used for translation and abstracting programs. This would suffice for minor early-warning services and relatively unsophisticated fact and document retrieval services.

2) "Semantic Machine." The "Semantic Machine" incorporates a process for resolving input ambiguity by reference to stored data, on the assumption that the true alternative is more probable. It probably represents significant improvement over EDGE for, for example, technical translation purposes, and would have value as an accessory on, for example, the input end of EDGE in increasing its tolerance of non-standard question forms.

3) Dialogue Analyzer. This would be given all the cues available to a human in dialogue via its own detectors or from an informant, including semantic and pragmatic items, and would have a

---

* These are merely particular instances of the general principle that when a new domain becomes relevant, the entity must distinguish and sufficiently understand that domain. (L.U.)
practical model of language-users. It could thus evaluate the probability of lying and other sources of error, etc., where they are relevant, and might be the minimum necessary equipment for efficient military intelligence source-evaluation work, lie-detecting in a courtroom, or analysis of psychiatric protocols.

4) Artificial Agent. Here, the artifact incorporates a self-consistent long-term goal system, full-scale learning routines, self-modeling and self-evaluation procedures and language, etc. The full-scale comprehension involved here might only rarely be essential for practical tasks, but the development of a number of artifacts capable of this would immediately raise the possibility that they could do independent, original work, which might be further improved by real dialogue between them, for which the above development would be necessary.*†

POSTSCRIPT

At the end of his classic paper, "Computing Machinery and Intelligence," Turing says:

* See also: Scriven, Ref. 23.
† We might equally well examine a number of additional combinations of basic functions, such as the following, that can be given a machine: 1) sense patterns; 2) store patterns internally; 3) "model" rather than store a rote representation of an external world; 4) output behavior; 5) move toward desired states (thus satisfying a need-value system); 6) output as a function of inputs; 7) output as a function of internal states; 8) output as a function of needs; 9) identify certain inputs as "talking about," hence, signs; 10) build up a grammatically structured language system of such signs; 11) identify the emitter of language utterances as an act or like it in one or more of the above respects; 12) identify certain inputs as coming from itself. (L.U.)
"We may hope that machines will eventually compete with men in all purely intellectual fields. But which are the best ones to start with? Even this is a difficult decision. Many people think that a very abstract activity, like the playing of chess, would be best. It can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English. This process could follow the normal teaching of a child. Things would be pointed out and named, etc. Again I do not know what the right answer is, but I think both approaches should be tried."

In this Memorandum, we have tried to indicate part of what is involved in providing a computer with the capacity "to understand and speak [a natural language, such as] English."
Appendix A

A NOTE ON TERMINOLOGY

D. M. MacKay

In the case of human beings, we do not say that brains think, feel, understand, etc., but that people do. When we come to consider artificial "brains" we meet the difficulty that no word exists for the artificial analog of "people." This can lead to confused debate unless terms are carefully defined and used.

We do have certain terms which embrace both the personal and the mechanical aspect of human beings. A "man," for example, may be said to think, etc., and also to weigh 150 lb and be made up of so many billion cells (though we would more often say the latter of his "body"). A "baby" is a borderline case where we are more inclined to use personal and physical categories indiscriminately.

In earlier papers, I have used the term "artifact" in this all-embracing sense (the artificial analog of "man" or "baby"), as distinct from "machine" or "computer," which I take to be more analogous to "brain" or "body." In doing so, I was not presuming to legislate on usage, but only indicating a distinction which needs to be made in some way if the comparison of natural and artificial information-processors is to be pursued intelligently. (1,8)

Because the term "artifact" can denote artificial constructs other than information-processors, I do not feel it to be ideal. "Automaton" would be better if it did not have the negative flavour of "mindlessness" or
the absence of personality. "Robot" is objectionable for the same reason. We need a term that neither implies nor precludes implicitly the fullest human-like characteristics.

The Gordian solution would be to invent yet another neologism, in (historically unjustified) hopes of keeping it free of misleading associations. Rather than make this cut, however, it seems worth considering whether the term "artificial agent" would not serve quite well and impartially for our purpose.

The term "android," which is shorter, has been used to denote an artificial human agent, and for this restricted purpose it would seem to be ideal (if its various associations could be lived down!). For the more general class of information-processors, however, a less question-begging term is still needed.

An "artificial agent" can be discussed (like a man) in terms of weight, bodily composition, mechanical principles, and the like, and also in terms of purposes, thoughts, understanding, and so forth, without presupposing any particular degree of competence at the personal level. With regard to the issues discussed here, the term "artificial agent" would seem to have the virtue of more genuine neutrality than any existing alternative.

In the present paper, we have tended to follow common usage in making "computer" serve as the omnibus analog of "man"; but, wherever questions would thus be intolerably begged, we have used terms like "computer system," "entity," "artifact," or "artificial agent" to preserve the logical distinctions analogous to that between a man and his brain.
Appendix B

DIMENSIONS OF UNDERSTANDING

M. Scriven

In this appendix are set out some of the specifications for an entity which can properly be said to understand. We do this by examining the dimensions along which a contrast is drawn between understanding and certain other conditions.

UNDERSTANDING THINGS VS. UNDERSTANDING LANGUAGE

One basic contrast appears to be that between, for example, "understanding engines" or "understanding horses" and "understanding the constitution" or "understanding French." In both cases, the behavioral criteria are the adequate performance of tasks from a certain family; but, the tasks are practical in one case, linguistic in the other. Someone who understands engines has to be able to locate the cause of and remedy stoppages with considerable, though not universal, success, has to be able to keep engines running better than the ignoramus, etc.; but he need not be verbally proficient at explaining how to prevent stoppages or describing the design difficulties in a particular model. The same kind of analysis applies to, for example, "understanding criminals" or "understanding children"; so, "understanding things" does not refer to things-by-contrast-with-people, only to things-by-contrast-with-language. Understanding language, or something expressed in a language, or understanding a language, typically
requires the capacity to answer questions, express the material in a different way, instruct others, etc., but it too may only require appropriate behavior, as with a child's or a dog's understanding of simple commands.

A dog may be said to understand the commands, "Lie down," "Come here," etc., on the basis of his obedience to them. It is important that a dog may even be said to understand them although he does not obey them, as, for example, when he would rather stay outside playing, he turns his head as we call, wags his tail, and rather sheepishly goes on chasing his friend. Understanding commands is not the same as responding appropriately; it is knowing how to respond appropriately. This is a condition of the knower which is dispositionally connected with the commanded behavior,* the connection being completed if the motivation to obey is stronger than the counter-attractors.

So understanding, whether of things (or people—or events or processes), or of language, is a capacity to perform certain tasks, the type of task varying with the subject understood.

UNDERSTANDING VS. IGNORANCE AND PARTIAL UNDERSTANDING

A major and obvious contrast, of course, is with the absence of understanding. It is of some importance to note the existence of degrees of understanding of both things

* A disposition is a state which manifests itself under certain conditions: it is a "state of readiness" or set. (M.S.)
and language, since some cognitive achievements are all-or-none, as we shall see.

UNDERSTANDING VS. KNOWING

One either knows three threes are nine or one doesn't. It isn't the kind of thing one is said to understand. One may know the date of one's birthday—but scarcely be said to understand it (though one may understand, in other senses, why it is when it is). To know a person is not the same as understanding him. To know a good place for a picnic makes sense, but it seems inappropriate to say you understand the place (of course, you do understand that there is a place).

Yet the proof of understanding, whether practical or theoretical, lies in demonstrations that one knows something.

The contrast seems to be that understanding is a term for knowledge of a number of related facts plus knowledge of their relationships. It is a molecular or configurational word, like the word "pattern" or "organization." One pin doesn't make a pattern, nor one man an organization, nor does a jumble of pins or a number of men. There is both a quantitative and a systematic requirement on knowledge if it is to be called understanding.

This applies easily enough to understanding engines or the theory of engines, but how does it apply to understanding a simple command?

UNDERSTANDING A MESSAGE VS. UNDERSTANDING A TOPIC

We may consider understanding things (where this is properly said) or understanding theories about them, to
be cases of understanding *topics*. Importantly different from this is understanding a message. The statement, "Computer power requirements are now less than 1 per cent of the pre-1955 level," can be understood by anyone who knows the meaning of the terms involved. It obviously does not follow that he understands the *topic* of computer power requirements, or even knows the single fact that solid-state technology is responsible for the change mentioned. Understanding a topic requires (at least) knowing a number of facts about the topic and their relationships. It is probable that topic comprehension does not require message comprehension, since a man may learn from experience as well as by being told.

When we are talking about message comprehension, what do we mean? There seem to be no sets of facts involved. But, we note that "understanding a message" is equivalent to knowing "what a message means"; i.e., knowing the conventional interpretation of the words when combined in this way; i.e., *knowing* a number of facts about the meaning of the words plus *their* relationships. The legitimacy of this unifying interpretation is supported if one thinks of the similarity between poring over a complicated sentence in a language one has not fully mastered and trying to understand the workings of a complicated mechanism. First, one recognizes or construes the meaning function of some parts; then one begins to develop long-chain hypotheses about their relationship; and finally, it all "falls into place"--i.e., one hits on a successful hypothesis. It is significant that we use the phrases "seeing what he means," "grasping the point," "getting the picture," with their
visual or tactual references to perceptions of a whole. It is an essential element in understanding, then, that it refers to a unifying (frequently a simplifying) conception of a complex phenomenon.

UNDERSTANDING AN UTTERANCE VS. UNDERSTANDING A LANGUAGE

VS. UNDERSTANDING WHAT A LANGUAGE IS

A dog may understand some utterances in English without understanding English. To say someone understands English is like saying he understands electronics: it does not guarantee he will understand every utterance in English, only a very large proportion of the most common types plus partial understanding of some more difficult ones. It is not at all to say he can formulate the rules of grammar. On the other hand, it seems plausible to say that no one could be said to understand a typical natural language if he did not realize it was a language. For a very simple language containing a few commands this may be debatable, as is the claim that this would be a language at all. But part of the behavioral repertoire for someone who is said to understand a natural language is the appropriate set of responses to terms like, "lying," "erring," "distorting," "exaggerating," "misstating," "language,"

*There may, however, be a fundamental asymmetry between understanding a topic and understanding a language. If a person understands a language, he has the capacity to acquire understanding of any topic. To understand a topic does not provide the capacity to understand other items, unless the topic is one of the fundamentals. (M.K.)
"saying," etc., which certainly requires the capacity to distinguish hearing a volcano erupt from hearing that it has, while realizing that both experiences under some circumstances give us some of the same information. In the ensuing section, we shall examine other requirements for understanding certain types of linguistic utterance, but this one has a special status in that it refers to the entire conception of understanding what language is.*

UNDERSTANDING AN UTTERANCE VS. UNDERSTANDING ITS CONSEQUENCES

People often hear what is said and understand it in an immediate sense, without assimilating it. The incorporation of new information into one's body of organized knowledge is close to the heart of "real" understanding of information material, and it requires several steps, of which recognition of the direct entailments or of consequences of input material is important. Propositions and sets of them have their own self-supported consequences, as well as joint consequences with other known propositions (next section).

Failure to recognize that "A is taller than B" informs you that B is shorter than A, counts immediately against utterance understanding. But, failure to see that Euclid's axioms entail Apollonius' Theorem does not undercut the claim that the axioms were understood. So, understanding an utterance implies understanding its self-supported consequences in some depth, but not in all depths.

*See also: MacKay, Ref. 3.
It is in this sense that computers frequently "understand" instructions better than those who give them.

UNDERSTANDING AN UTTERANCE VS. UNDERSTANDING ITS SIGNIFICANCE

The significance of a proposition involves the joint consequences of a proposition and other known propositions, and the consequences of a proposition (and deductions from it) for other known propositions. Sometimes we say this is what a proposition "really" means, or "implies," or "tells us." If I say that a particular car has an aluminum engine block, you may (a) understand this message and certain deductions from it, (b) understand its consequences in terms of better weight distribution and, hence, ride, tire wear, and handling. The second would be what the proposition meant for an expert; but we must not legislate that such a rich meaning must be seen in the proposition by everyone who can be said to understand it.

UNDERSTANDING THE SIGNIFICANCE OF AN UTTERANCE VS. UNDERSTANDING A TOPIC

One cannot understand the significance of an utterance in the sense just described without having some acquaintance with some of the local factual geography. But it does not presuppose knowledge of the whole field; and understanding of a topic develops outward from understanding utterances and coming to see more and more of their consequences.
SIGNIFICANCE OF AN UTTERANCE VS. SIGNIFICANCE OF THE
CIRCUMSTANCES OF UTTERANCE

We sometimes refer to an entirely different dimension from those mentioned so far when we use the terms "significance" or "meaning" of an utterance. We are referring to our inferences from the fact that the utterance is made at all, that it is of a certain kind, that it is made in a certain tone, accompanied by certain gestures, comes at a certain time, and so on. We might call this circumstantial information by contrast with content information. In problems of machine translation of verbal output, for example, capacity to handle such circumstantial cues would probably be essential in reducing the ambiguity of natural languages and would contribute an order of magnitude more information than the use of "semantic machines."

ADEQUACY, DEPTH, AND RANGE OF UNDERSTANDING

Understanding is always relative to a certain level of interest. Someone who can answer all questions about the nature and behavior of gases to a certain degree of precision may be correctly said to have a very good understanding of the subject. There would still be great stores of information of which he is unaware and which would be known to someone who could be said to have an even more thorough understanding. One must not suppose that the existence of micrometers shows that remarks like, "He is 6'1" tall," are inaccurate in any useful sense. Understanding is entirely satisfactory if it handles the questions for which it is needed. Adequate understanding does
not require any predetermined depth of understanding; it requires only whatever depth is required.

Early in our discussion of synthesizing comprehension we encountered the feeling that range of understanding is also crucial: It is not enough that an entity answer all questions about a particular narrow field, such as chess or family relationships, with great efficiency; it also seems to be necessary for it to relate this field to others in numerous ways. This is partly the demand for understanding significance and not just message-content. Partly, it springs from the feeling that the ramifications of any part of natural languages into the whole multi-modal range of human experience must be captured by any entity that may be said to understand natural language. Language is a phenomenon of social evolution in a complex environment, and--it might be said--cannot be properly understood by an entity lacking understanding of this social function of language.

UNDERSTANDING VS. MEMORIZING ("HAVING" A MODEL VS.
INCORPORATING A MODEL)

The purpose of essay exams, and good exams in general, is to separate the memorizers from the comprehenders, as well as both from the ignorant. We have already stressed the complex and integrative features of understanding. Now we come to the more elusive feeling that understanding must be distinguished from mere ability to answer questions (which might have been memorized). A computer with enormous storage capacity might gradually become able to answer 99.9 per cent of all questions about protein chemistry that
are asked of it, including those which call for "knowledge" of the relation of the facts and theories it can produce. It is just a semi-automatic, highly redundant encyclopedia. For practical purposes, it is better to have a more efficient storage system; i.e., an associative hierarchy or an analog. (12,13) The use of such design improvements does not necessarily increase the understanding but only the efficiency of the machine. It might be said to incorporate a model but not to have a model. But, it is a step towards understanding and away from rote memory procedures, because it is a step nearer to integrating and simplifying the stored data. The "memory" is not quite so much a matter of "rote," but is still just information storage. In particular, such machines have no knowledge that the organization of their memory is based on a model, hence they do not have a model in the crucial sense of having what one knows is only a symbolic representation of reality. Comprehension of a topic, or of language in general, requires recognition of the distinction between model and reality or between language and reality. * But understanding a language of very simple commands is not different from understanding growls from a lion and needs no such sophistications. The depth shrinks, and with it the distinction between understanding and mere reflex responses. As extra modalities or informants or topics are added, the range required increases, and the problem of relating them becomes

*Programs that modify their internal store as a function of experiences with data already make this distinction. (L.U.)
more pressing. In domains of sufficient complexity, understanding is sharply different from reflex responses.

**UNDERSTANDING VS. MISUNDERSTANDING**

The most sophisticated contrast of all is that between understanding and misunderstanding, since the latter concept cannot be understood without the former. Misunderstanding is, subjectively, understanding, and, objectively, a mismatched model.*

**UNDERSTANDING VS. THE FEELING OF UNDERSTANDING**

The feeling of having understanding is only an unreliable symptom of the true condition. It is not hopelessly unreliable since it is a learned response and would be extinguished if never reinforced. From the computer designer's point of view, the equivalent of the "Aha!" feeling in understanding is a report to the central organizer of a completed filing routine. It can be set off in several ways by feedback; e.g., when:

a) Filing has proceeded smoothly to completion;

b) Apparent incompatibilities have been resolved;

c) A standing tagged question has been answered by the input;

d) Trial runs through the associated material reveal that the model can answer certain general questions;

e) The input fits exactly into an existing structure which is known to be a satisfactory storing system;

f) The input effects a link which enables a great condensation of other stored material to be effected.

*See §3 of Appendix C.*
Appendix C
COMPREHENSION OF UTTERANCES
D. M. MacKay

1. Utterances are exchanged between agents as a means of: a) initiating, modifying, or confirming one another's action (overt or internal) or conditional readiness for action; b) evoking emotional or other experiences. The target of an utterance may be: a) the recipient's "map" of the world (i.e., what he reckons with as fact or possibility in planning and executing action); b) the recipient's normative system which determines the conditional priorities of various norms and goals, and actions directed thereto; c) the recipient's (or just the originator's) sensory-aesthetic or other systems not directly concerned in the organization of action adaptive to the external world (e.g., certain aspects of poetic utterance, etc.).

2. The meaning of an utterance (intended, standard, or received) can be defined as its selective function (intended, standard, or actual) on the range of possible states of the appropriate system.\(^{(4-7)}\) Thus, meaning is defined as a relationship, not an absolute property. An utterance is meaningless if it has no selective function (for one reason or another) on the appropriate range. We have to distinguish between meaningfulness due to: a) lack of definition of selective function (e.g., nonsense syllables); b) absence of the appropriate range in recipient (e.g., "colour" to a blind man); c) incompatibility
of two or more components of the selective operator (e.g., "the milk is isosceles").

Clearly, utterances can be partially meaningful (or meaningless) under the above heads.

3. In this framework, misunderstanding can be defined as discrepancy between intended or standard and received selective functions. Note, however, that this is the passive sense of misunderstanding (e.g., "a misunderstanding exists") rather than the active sense (e.g., "you are misunderstanding me"). To misunderstand is to perceive an utterance as calling for an adjustment of organization discrepant with that intended or purported.

4. The last words emphasize that the effort to understand an utterance is an effort to match the recipient in some sense to the originator--i.e., it presupposes an intention in the originator and an awareness (explicit or implicit) of this by the recipient. We must beware of any behavioural tests of understanding which omit to test for this.

5. An utterance as understood by a recipient has two distinct aspects--its form and its weight.

The first determines the form that the recipient's organizing system would assume if the utterance were fully accepted--i.e., if the selective operator it defines were fully applied to the organizing system.

The second determines the actual degree of coupling between the internal representation of this selective operator in the recipient, and his organizing system.

We thus picture the receipt of an utterance as a two-stage process, somewhat like the construction of a patch-
board followed by the decision to plug it in—but different in that the final coupling need not be all-or-none.

6. In these terms, we can briefly characterize the understanding of a variety of utterances.

6.1. A statement is aimed at the recipient's map-making system. Its received weight (the extent to which it is allowed to determine the form of the map) depends on the recipient's evaluation of the originator's goal in uttering it and of his reliability as a source (e.g., he may ask, "Is he trying to deceive me?"—"Does he have access to necessary data?").

6.2. A command is aimed primarily at the recipient's normative system. It is calculated to secure action of a given form by altering the goal-priorities, rather than by mere physical reflex stimulation, of the recipient. Its perception as a command requires recognition that the recipient is in principle under feedback from the originator, with equilibrium dependent on his taking the required action. Its weight depends on the extent to which disequilibrium here is negatively valued (i.e., to ask, "What if I were to disobey?" would be relevant).

6.3. A request differs from a command in relying for its weight upon the positive value (to the recipient) of satisfying the originator, rather than evoking mere negative avoidance of disequilibrium with him.

6.4. A question is aimed formally* at updating the originator's own organizing system, by way of the

*Examination questions merely require the recipient to answer as if the examiner needed to be updated.
recipient. Questions may be regarded as requests (or commands according to their form or tone) to construct a linguistic tool with the specified updating function. Their weight is thus derived in an analogous manner to that of requests and commands, with the added consideration that for many people the updating of others is a positively-valued activity in any case.

6.5. An instruction differs from a command in specifying not an action or a goal, but a program, whose goal may be only conditionally named if at all (e.g., "To reach the house, drive five miles north on route 1, then two miles east on 50").

6.6. A warning is aimed at the recipient's normative system by way of his map. It presupposes a coupling (usually negative) between the internal representation of what the warning indicates and the evaluatory system which determines goal-priorities (e.g., an air raid warning). Its effective weight depends both on the strength of this coupling and on the recipient's evaluation of the motivation and reliability of the originator.

6.7. Advice, in general, may include instructions and warnings. It is characterized, however, by the implicit or explicit conditional, "If I were in your place ...." The weight claimed by advice thus depends on the extent to which its originator believes himself to have adopted the recipient's goal-structure, etc., when constructing it. The weight actually given it by the recipient will then obviously be calculated from his evidence on this point, as well as from the intelligence, knowledge, etc., he attributes to the advisor.
6.8. A promise is aimed at the recipient's map of the originator's goal-settings. It indicates not only that the originator shares the recipient's goal named in the promise, but also that he recognizes himself in principle to be under feedback to the recipient in this connection (as in the case of commands).

7. As argued in Ref. 6, all utterances also have an indicative function, whether or not they are conventionally aimed at the recipient's cognitive map. Thus even a question or a command implicitly indicates certain presuppositions on the part of their originator. Understanding of these indicative aspects may be symptomatic or linguistic or both, and need not coincide with understanding of the conventional (interrogative, imperative, or other) aspect. A question such as "How long have you lived in Mexico?", for example (addressed to someone who never has), can be indicatively understood and replied to appropiately, even though it is interrogatively nonsensical.
Appendix D

ON THE POSSIBILITY OF ENTITIES WHICH CAN BE SAID TO UNDERSTAND

M. Kochen and L. Uhr

INTRODUCTION

Under what conditions does the existence of an entity capable of powerful intellect and genuine understanding seem at least theoretically possible? Can such an entity be designed the way machines are designed, or must it be "nurtured" or grown into existence? If the latter, what are the minimal capabilities with which it must be endowed when first constructed? What principles for acquiring what has not been explicitly supplied by the designer must be provided? How must it contact what sort of world?

SKETCH OF TENTATIVE BASIC MODEL FOR COGNITION

At least two types of entities, both modeled as finite-state automata, constitute the basic model. Each is characterized by the following:

a) There are a number of states, some stable, with a tendency toward stable states.

b) There are continual disturbances from the outside causing deviations from stable states.

c) New energy is needed to maintain such motion due to energy degradation.

One type of entity, denoted by M, represents the machine supposed capable of acquiring understanding through interaction with its environment. The other type of entity, denoted by E, represents M's environment excluding M.
M differs from E in that E is immensely more complex. Thus, state transitions in E cause significant disturbances for M, but not vice versa. This means that E could generate output sequences that M could not analyze (i.e., reconstruct and predict, based on the principle that generated them and not by imitation); E could generate sequences of greater complexity than M could generate; E may contain entities of complexity comparable to that of M. Two M's are of comparable complexity, exchange little energy, and are relatively undisturbed by one another. Their stable equilibrium states are such that their attainment may well depend upon cooperation.

Machine M consists of an input tape, output tape, internal state, and self-modifying apparatus such that its current state and output state is co-determined by its state and input at the previous time epoch. The following more detailed machine operations are postulated for M:

a) Shift. This moves input tape and output tape.

b) Test two items for equality. This matches two items, outputting 0 to 1 for match or no match.

c) Store an item in a specified cell. This punches onto output tape if the specified cell is on the output tape; or stores in memory, changing the internal state, if the specified cell corresponds to an internal state variable.

d) Compose two items to form a third.

e) Associate two items relative to a specified type of association (relation).

f) Include one item in a class.

An M inputs data one "moment" (dt) at a time. A moment's data might be a single letter from a primitive alphabet (e.g., 0,1) or an n-dimensional array of such letters.
An M is able to grow its store in the sense that it can add nodes and edges to its memory* (which begins empty) as it needs to.

An M has a set of "learning rules" that determine what and when it should add to its memory graph as a function of the graph's present state and input patterns.

An M has a set of "behaving rules" that determine what and when it should output as a function of its memory and inputs.

An M has a set of subroutines that guide the flow of information from input to memory graph to output.

An E is an M with many states, a relatively small and simple need-value system, large output and input capability, and simple functional relations relative to M.

Essentially, we can think of an M as being in communication with its environment E which is, from M's point of view, another M. In general, E will be very large with respect to M, but will itself compute the functions y and s, as M does. The situation of dialogue between M-like entities arises when there is more than one M.

An M accepts inputs that have been put on its tape by E. It "recognizes" patterns in these inputs by shifting and matching (testing for equality), essentially seeing whether these are patterns that it has previously decided were worth recognizing again. These will be patterns that M has inferred co-occur with need reduction (i.e., with

*The memory structure is modeled as a graph in which the nodes are things, compounds of things, or classes, and the edges are relations.
movement toward a preferred state). M will also be outputting patterns that its internal state suggests are appropriate, given the present need-state and environment to put appropriate patterns of value on its input tape. M will also be reorganizing its internal state as a function of inputs, following rules for making inferences that appear appropriate for getting more valuable inputs onto its tape by E in the future.

In general, M builds up a memory graph that composes from primitive into higher-order things (e.g., primitively discriminables to strokes to letters to words to idioms), associates things with other things, and makes things members of other things, such that characteristics of a single thing are generalized to any member of the class.

EXAMPLES OF ALGORITHMS

A. Watch input sequence for recurrent patterns, and, using the "compose" operation, store these in memory. For example, treat E as a sequence generator and try to build up a minimal description of the generator; or, store in memory sequences that recur.

B. Store co-occurrences of patterns defined in memory by algorithm using the "associate" operation, along with A, with frequency of co-occurrence. This might also stipulate that one of the two patterns must also be a need-reducer.

C. Restructure (classify) memory. For example, when two patterns co-occur with a third, tentatively group these two into a class and test whether this generalization holds.
D. Learn to imitate (given two patterns, α and β, on the input tape, produce β on the output tape when α appears on the input tape).

E. Discover input patterns "caused" by its own output patterns. For example, store co-occurrences of input patterns and output patterns, and test hypotheses thus formed by running experiments in which M outputs patterns to see whether the expected pattern will appear on its input tape.

Other algorithms are needed for functions such as the following:

F. Store information as to type of co-occurrence as defined by B.

G. Learn transforms.

H. Learn to handle recursive series.

I. Weigh stored patterns by costs and modify values.

CREATION AND GROWTH

The situation in which natural cognitive entities arose is clear. In a relatively extensive environment there arose at least one organism with capability to no longer merely react to physical stimuli, but to respond, to act appropriately to these stimuli. Thus, such an organism gained from its responses, in some sense making the environment--the small part of the universe with which it was capable of interacting--more benign and rich in things it valued. To what extent might such a situation be necessary for understanding?

Life could not have arisen in a small universe. To the extent that the minimum necessary organism is complex,
and the universe is unstructured, we would need at the least a universe that was large enough or had gone on long enough so that the minimal organism could arise somewhere. The size of the organism relative to the universe would be overwhelmingly small (depending upon actual diversity of the universe and necessary complexity of the organism)—small enough so that such an unusual occurrence happened at least once. Given such a beginning of life, the organism to which we may want to apply the phrase "it understands" starts, continues, and possibly remains eternally in a position of overwhelming ignorance relative to understanding. It understands only a very small part of a very large universe.

From a predicament of this sort, several interesting hypotheses seem likely. The organism must be capable of adapting—"evolving" through the species and "learning" through the individual. But this is adaptation of a very unusual sort. The organism must adapt to the universe (as presented to it in its immediate environment) in order to continue and to accelerate the acquisition of experience of value. Thus, the organism must have some raison d'être—
a) some need-value system such that there are states of its immediate environment that are valued by it; b) some sensing system that will allow it to be aware of these states; c) some effector system—some ability to act upon this environment in such things. That is, the notion of a small organism that arises in a large universe because it can benefit from certain relatively very small and common parts of that universe entails the necessity of a sensing system, a need-value system, and a response system.
The essential enterprise of such a system is to come to know better, in order better to cope with, its universe. Thus, the more it can sense, the more it can manipulate; and the more it can get, the better off it will be. So we would expect it to develop (if it could) better sensors, better decision-makers as to what to do as a function of particular sensed world-states, and better effectors. We might think of such a system as an experimenter, attempting to form hypotheses, adduce evidence, and build theories about the world, theories to guide action that will lead to fruitful consequences. This, we assert, is what perception and learning are: the attempts of a fallible and weak system for discovery and induction on a universe sufficiently large to be completely undiscernible and sufficiently benign to be partially discoverable almost always.

Given such a situation, one thing that will certainly arise, if possible, is the development of additional such systems, and the development of means of communication between them. There is benefit in the existence of several such systems rather than one, since more means more experiments going, more information being gathered about the world. And there seems an advantage to having this go on in at least slightly different parts of the universe—far enough apart so that different parts of the universe are being sampled, but near enough together so that experience can be pooled—that is, if there is some method for pooling. But it seems safe to assume that the sensing and acting systems that have already emerged are sufficient basis for communication systems. What we are pointing out
is that there will be a need for others, and for communication with others, simply because there is the general need for growth and this is one powerful and possible method for growth.

UNDERSTANDING

Can we conceive of some static state—"understanding"—as it exists at a given moment, divorced from how it arose or where it will go? This is a common use of the term in everyday language, and would seem to have meaning. This is the sense in which we say, "He understands that when you push that button the lights go on," or, "He understands that bishops move diagonally on the chessboard," or, "He understands the sentence, 'It is raining.'" "Understanding" in this sense seems easy to give to a computer—dangerously, suspiciously easy. "Understanding" in this sense includes examples too homely for acceptance (e.g., "The door understands the push") and of as great an amount and as different a kind of complexity as we would wish (e.g., "He understands that he understands that I understand his intentions in making the statement, 'I don't understand her.'"). The differences seem to be results of differences in complexity of domain, interrelation between parts, and abstractness.

We can choose to use the word "understanding" for such states if we wish. But we observe the peculiar situation that all bona fide understanders—i.e., all the higher animals—are also understanders of the other, dynamic sort, capable of coming to know. We suggest that there may be some essential relation between the two; we assert that
the two are necessary for both living organisms in their evolutionary predicament and any entity that we might wish to beget and get to understand. We will certainly never design* and build an understander of any interesting domain; what we must do is try to establish some sufficient entity in some sufficiently benign environment, worry a bit like parents, and hope for the best. This, then, seems necessary for intelligence. It is also necessary for understanding in the dynamic sense. Is dynamic understanding in fact, or of necessity, necessary for static understanding? Thus a universe not yet exhausted by the understander is necessary for understanding even of those substructures of the universe that are understood.

Such a line of argument suggests that we must not carve out a domain for a system to comprehend. We cannot use a cookie cutter on the universe. Rather, we must leave it in its inexhaustible connectivity and give the proto-understander a window onto his part of the world. A window allows him to see where he will.

CONCLUSIONS AND SUMMATION

Understanding is assumed to arise from the necessity of a sufficiently complex and sensitive entity to assure its continued existence in an environment of immensely

*The term design is here meant to denote complete specification for construction, possibly also training, implying a definite training sequence in controlled environment (e.g., a Skinner box). All possible exigencies in the environment are anticipated and appropriate responses are a priori specified.
greater complexity than itself. Though the term "understanding" is commonly used to denote many functions and states, its use to denote the capability of making leaps towards understanding of anything in an unspecified domain of discourse seems to us the most significant.

To achieve "understanders" and "intelligent" entities, we cannot expect to analyze domains and then design appropriate machines. Rather we must try to get a grasp on methods for learning, discovery, and introduction, and design machines that are general learners. These machines must be effectively coupled to sufficiently interesting environments. By "sufficiently interesting" we mean that the environment must be capable of transmitting the information we want the machine to absorb. Environments will frequently contain other machines, in which case language might be expected to arise as a vehicle for sharing between machines whatever knowledge each individual has amassed, and for arriving at community of interest and action.

Some simple aspects of the design of such a "cognitive" agent appear to emerge. To be connected with its environment, it must sense and act. To have any reason for absorbing information, for becoming "cognitive" in the first place, assuming a need-value system with a value-maximization principle or something equivalent, seems plausible. From this, it must organize functions that can test to see whether desirable patterns, and indicants of desirable patterns, are there, and then it must act appropriately--e.g., by placing new pattern-configurations in its memory or on its output tape. Thus it must have as basics such
abilities as "match," "shift," "copy," and such connectors as "compose," "associate," "include."

We hope that now it might be sensible to ask: What is the minimum complexity for such an entity that will allow it to try to become an understander of a universe complex enough to contain it?
REFERENCES


