Communication with Alien Intelligence
Marvin Minsky

In memoriam: Hans Freudenthal

Published in *Extraterrestrials: Science and Alien Intelligence* (Edward Regis, Ed.) Cambridge University Press 1985. Also published in Byte Magazine, April 1985.

When first we meet those aliens in outer space, will we and they be able to converse? I'll try to show that, yes, we will–provided they are motivated to cooperate–because we'll both think similar ways. My arguments for this are very weak but let's pretend, for brevity, that things are clearer than they are. I'll propose two reasons why aliens will think like us, in spite of different origins. All problem-solvers, intelligent or not, are subject to the same ultimate constraints–limitations on space, time, and materials. In order for animals to evolve powerful ways to deal with such constraints, they must have ways to represent the situations they face, and they must have processes for manipulating those representations.

ECONOMICS: Every intelligence must develop symbol-systems for representing objects, causes and goals, and for formulating and remembering the procedures it develops for achieving those goals.

SPARSENESS: Every evolving intelligence will eventually encounter certain very special ideas–e.g., about arithmetic, causal reasoning, and economics–because these particular ideas are very much simpler than other ideas with similar uses.

The "economics" argument is based on the fact that the power of a mind depends on how it manages available resources. The concept of *Thing* is indispensable for managing the resources of Space and the available substances, which fill it. The concept of Goal is indispensable for managing the ways we use our Time–both in regard to what we do and what we think about. Aliens will use these notions too, because they're easy to evolve and seem to have no easily evolved alternatives.

The "sparseness" theory makes this thesis more precise. It holds that almost every evolutionary search for ideas will encounter certain common ideas. These are those peculiar concepts for which there are simply no easily accessible alternatives, that is, other very different ideas that can serve the same purposes. This is because, I'll argue, certain ideas are islands by itself, because there is nothing which resembles them that is not either identical or vastly more complicated. I will only discuss the example of arithmetic, but I suspect that there are other ideas about Objects, Causes and Goals that have the same character.

CRITIC: What if those aliens have evolved so far beyond us that their concerns are unintelligible to us, and their technologies and conceptions have become entirely different from ours?

Then communication may be infeasible. Our arguments may apply only to those aspects of intellectual efficiency which constrain relatively early stages of mental evolution–stages in which beings are still concerned with survival, communication, and expansion of control over the physical world. Beyond that, we may be unable to sympathize with what they come to regard as important. Yet still, we can hope to communicate with whatever remains of the mental mechanisms they use to keep account of space and time consumed, for these may remain as sorts of universal currency.

CRITIC: How can we be sure that things like plants and stones, or storms and streams are not intelligent, in other ways?

If one can't say how their intelligence is similar, it makes no sense to use the same word. They don't seem to solve the kinds of problems we regard as needing intelligence.

CRITIC: What's so special about solving problems? Anyway, please define "intelligence" precisely, so that we'll know what we are discussing.

No. It's not our concern to tell people how to use a word. We only want to discuss communicating with intelligent aliens. Let's just use "intelligence" to mean what people usually mean, namely, the ability to solve hard problems – like the kinds a species must solve to build such things as spaceships and long-distance communication systems.

CRITIC: Then, you should at least define what a "hard" problem is. Otherwise you may unwittingly become a human-mind chauvinist. For instance, we assume it took intelligence to build the Pyramids – yet coral reef animals do things on even larger scales. Would you claim that we should therefore be able to communicate with them?

No, because, while humans solve such problems, it is only an illusion that the coral animals do. Speed is what distinguishes intelligence. No bird discovers how to fly: evolution used a trillion bird-years to 'discover' that–where merely hundreds of person-years sufficed. And where a person might take several years to find a way to build an oriole's nest or beaver's dam–no oriole or beaver could ever learn such things at all, without the ancient nest-machines their genes construct inside their brains. But that is not intelligence: no such machinery seems capable of solving wide ranges of new, different kinds of problems. It would make sense to try to speak to any animal that learns to solve new, hard problems–but not problems which we work trillions of times faster. What makes us able to do such hard things so fast? The following ingredients seem so essential that we can expect intelligent aliens to use them, too.

SUBGOALS ----------- to break hard problems into simpler ones.
SUB-OBJECTS ------ to make descriptions based on parts and relations.
CAUSE-SYMBOLS--- to explain and understand how things change.
MEMORIES ---------- to accumulate experience about similar problems.
ECONOMICS -------- to efficiently allocate scarce resources.
PLANNING ----------- to organize work, before filling in details.
SELF-AWARENESS-- to provide for the problem-solver's own welfare.

But aren't these an arbitrary few, from myriads of still unknown other possibilities? Why can't the aliens do such things in other ways? I'll argue that these problem-solving schemes are not as arbitrary as they seem.

THE SPARSENESS PRINCIPLE

Why does it seem so clear to us that Two plus Two must equal Four? Such mysteries have long concerned philosophers–to say why certain concepts seem to come into our minds as though they need no prior experience or evidence. I claim that it's in part because of this computational phenomenon:

The Sparseness Principle: When relatively simple processes produce two similar things, those things will tend to be identical!

If this is so, then certain "a priori" ideas will appear as a natural consequence of the way a mind evolves by selection from a universe of possible processes. This would also explain could why different people can communicate so perfectly about such matters as arithmetic, although their minds differ in other ways. And so, it may apply to aliens, too. I will explain the sparseness principle by recounting two anecdotes. One involves a technical experiment, the other, a real-life experience.

A TECHNICAL EXPERIMENT. I once set out to explore the behaviors of all possible processes–that is, of all possible computers and their programs. There is an easy way to do that: one just writes down, one by one, all finite sets of rules in the form which Alan Turing described in 1936. Today, these are called "Turing machines." Naturally, I didn't get very far, because the variety of such processes grows exponentially with the number of rules in each set. What I found, with the help of my student Daniel Bobrow, was that the first few thousand such machines showed just a few distinct kinds of behaviors. Some of them just stopped. Many just erased their input data. Most quickly got trapped in circles, repeating the same steps over again. And every one of the remaining few that did anything interesting at all did the same thing. Each of them performed the same sort of "counting" operation: to increase by one the length of a string of symbols–and to keep repeating that. In honor of their ability to do what resembles a fragment of simple arithmetic, let's call these them "A-Machines." Such a search will expose some sort of "universe of structures" that grows and grows. For our combinations of Turing machine rules, that universe seems to look something like this:

```
               X X
             X  X  X
            X X   X
           A X   X  A X
         X   X  A    \
        X X   X   X X  A
      X    X  X  A  X  / \
       X A    X   X XX/XX \ XX
     X X /X X  X XX X /X XXX\ X
    X X A X  X  X X  A  X X  A  X
   X AX X X X X X XXX   XX X / \ XXXX
   XX A XX XXXX XXX  X X    A  X A X X
  X XXXAXXX XXX XXX XXX  X / \XX XXX XXX
 X XX XXX XXX XXX XXXX B XX XXX XXX XXX X
X X XXX XXX XXX XXX XXXX/ \XX XXX XXX
```

The 'X's are useless processes that don't do anything at all. The 'A's are those little "counting machines." In effect, they're all identical! These A-machines must be the early seeds of our other ideas about arithmetic–and it seems inevitable that somewhere in a growing mind some A-machines must come to be. What of other, different ways to count? Much, much later will appear some B-machines–which act in ways that are similar, but not identical. However, our experiment suggests that even the simplest B-machine must be so much more complicated that it is unlikely any brain would imagine any B-machine before it first found many A-machines.

In some sense, this little thought-experiment resembles an abstract version of those first experiments in which Stanley Miller and Harold Urey set out to explore, with real chemicals, the simplest combinations of constituents.

They started with a few elements like Hydrogen, Oxygen, Nitrogen, Carbon, and Phosphorus and found that those chemicals react first to make simple molecules and then went on to form peptides, sugars, nucleotides and what-not. Of course, we would have to wait much, much, longer before the appearance of tigers, woodpeckers, or Andromedans.

A REAL-LIFE EPISODE. Once, when still a child in school, I heard that *minus times minus is plus.* How strange it seemed that two such negatives could "cancel out"–as though two wrongs might make a right, or like that self-refuting falsehood which declares itself a lie. I wondered if there could other, different thing, still like arithmetic, except for having yet another "sign". Why not make number-things, I fantasized, which go three ways instead of two? I searched for days, making up new little multiplication tables. Alas, each ended either with wrong arithmetic (like making One and Two the same) or something with no signs at all. I gave up, eventually. If I had persisted, like Gauss, I might have encountered the complex numbers–which exist, but have essentially *four* signs or, maybe, Pauli's spin matrices–which also have four signs. But no one ever finds a three-signed version of arithmetic–because, it seems, they simply don't exist.

Try, yourself, to make a new number system that's like the ordinary one, except that it "skips" some number, say, 4. It just won't work, because everything will go wrong. For example, you'll have to do something about "2 and 2". If you say that this is 5, then 5 now has to be an even number, and so must 7 and 9. And then, what's 5 plus 5? Should it be 8, or 8, or 10? You'll find that you have to change all the other numbers' properties, to make the new system be at all like arithmetic. And when you're done, you find you've only changed those numbers' names and not their properties at all.

Now make two different numbers be the same–say, 139 and 145. Then 6 must be zero and 4 plus 5 must be 3. A little change–but now you find that the sum of two numbers can be smaller than either. Such systems have a certain usefulness in abstract mathematics–but are utterly useless for keeping track of real things. And so it goes.

There simply is no any way to take one number out or put another in–nor can you change a single product, prime, or sum. What gives Arithmetic this stark and singular rigidity? Why can't we make the smallest hole in it, or make it stretch or bend the slightest bit? The whole thing stands "there" stiffly–you have to take it, all or none–because it's isolated as an island in that universe of processes. That self-same A-machine exists, immutably complete, as part of every other process which can generate an endless chain of different things.

I wonder if it's dangerous to make our children think so much about arithmetic–if, when it's seen this way, it leads to such a singularly barren world. Some of us discover in it a universe of different ways to add, and different ways to think up more such ways. But most children find it dull–just endless, pointless pain and rote–the tedium of working abstract clay too cold and stiff to mold and shape. The only ones who benefit are those who, seeing that they cannot bend the rules, distort instead the ways they're used.

Both anecdotes appear to show that any entity that searches through the simplest processes will soon find fragments which don't just resemble arithmetic; they are arithmetic exactly. It is not because of our own lack of inventiveness or imagination, but is just a fact about the computational geography of a world more rigid and constrained than any "real" thing we know.

THESIS: All processes or formalisms that resemble arithmetic are identical to arithmetic, or else unthinkably complicated and arbitrary. This is why we can communicate perfectly about numbers.

What has that to do with aliens? They too must evolve–which means developing by searching through some universe of possible structures. All evolutions, too, must tend first to examine combinations of the relatively simple systems that they first select; all others will be impracticably inefficient.

Why do ideas occur as isolated islands–with nothing similar nearby–in these worlds of evolutionary search? It is hard to make the reason perfectly precise, because the concept of two "similar" ideas depends on what you want to use them for. Still, the principle is clear: small sets of rules can generate vast worlds of implications and consequences. But–most large worlds cannot be made from smaller sets of rules–because there simply aren't enough little sets to go around! That's why we can't go back again, to make a little change in a world of consequences, and hope to see a similarly small change in its rules. Then, the simpler the sets of rules, the fewer such that there can be–hence the more rigid must be their worlds of consequence. There just can't be much flexibility or continuity in the earliest phases of an evolution.

CAUSES and CLAUSES

If alien minds were entirely different from ours, communication might be impossible. That would happen if the way we think were just an evolutionary accident. But though each evolution is composed entirely of accidents, that only holds for fine details. On larger scales, each evolution tends to first try relatively simple ways at every stage. So, since we're first on earth to grow a large intelligence, we probably did it in some likely way. This also must have shaped the ways that we communicate. I'll explain the idea in a form so strong that at first it will seem preposterous–in terms of the grammar of human languages. I'd rather talk in terms of how our thinking works, except that we don't understand this well enough yet.

For every difference, move, or change, our language-syntax makes us seek some cause. No matter that no actor's on the scene: we'll find one, real or fantastic. That's why we say, "It soon will start to rain". What makes us postulate a cause, no matter if we're right or wrong? I claim this isn't merely surface form, but stems from deeper causes in the ways we think. My guess is that we have developed special brain-machinery to represent objects, differences, and causes–and much of our thinking is based on using mental symbols in these ways:

OBJECT-SYMBOLS represent things, ideas, or processes. In languages, they often correspond to Nouns. Our minds describe each scene, real or mental, in terms of separate object-things and relations between them.

DIFFERENCE-SYMBOLS represent differences between, or changes in OBJECTS. In languages, they correspond to Verbs. When any object undergoes a change, or two objects are considered at once, the mind ascribes some DIFFERENCES.

CAUSE-SYMBOLS. When any DIFFERENCE is conceived, the mind is made to find a CAUSE for it–something to be held responsible. We use a clever mental trick of representing causes in same way that we represent objects.

CLAUSE-STRUCTURES. Whatever we can express or describe, we can treat its expression or description as though it was a single component inside another description. In languages, this corresponds to using embedded phrases and clauses.

That final trick–of representing prior thoughts as things, gives our minds the awesome power to use the same brain-machinery over and over again–to replace entire conceptualizations by compact symbols, and hence to build gigantic structures of ideas the way our children build great bridges and towers from simple separate blocks. It lets us build new ideas from old ones; in short, it makes it possible to think. The same in our computers, too.

This must be why our languages use structures that can be re-used: our thoughts themselves must use the same machinery repeatedly to reach unlimited variety. Unless our aliens do that too, they can't turn thoughts upon the products of their thoughts–and won't have general intelligence–however excellent their other rigid repertoires of skills may be.

CRITIC: You might as well argue that the aliens would speak English, if you claim they, too, use nouns and verbs and sentences. But what if they don't think in terms of objects and actions at all?

I don't think it's an accident, the way we think in terms of thing and cause. It forces us to always wonder who or what's responsible, whatever happens. I claim it is the best way evolution's found to make us search to find dependencies that help to predict–and hence to control–not only the world outside but also things that happen in the mind. I think it's also why we all grow up believing in a Self: that "I"–in "I just had a good idea"–stems from that same machinery. For, since you are compelled to find a something to explain the things you do–why, then, that something needs a name; you call it "I."

CRITIC: Why can't those aliens perceive entire scenes as wholes instead of breaking them down into our clumsy things with properties? They might instead see what there really is, holistically, as steady flow of formless space in time, instead of arbitrary separate mind-made fragments of approximations to reality?

We always yearn for better ways–and that's a healthy tendency. But worshipping holistic schemes can blind us to the power we gain from usual ways of separating things. Each animal must pay some price, in clumsiness and nourishment, for each computer carried in its brain. That trick of seeing 'things' is what allows our minds to use the same machine for many different things–just as clause-structure in language lets one focus the entire mind on each small part of a description. Non-holistic methods factor situations into parts–to let us apply our whole mind-machine to each part of a problem.

Enthusiasts of holism just never seem to see the price of "seeing everything at once". There have been some speculations that brains might use something like holograms for memories, but there is little basis for such ideas. For one thing, for a given investment, holograms store no more information than other methods. Now it is true that they facilitate certain kinds of recognitions, for example, to decide whether a certain picture contains a copy of some certain other picture. But the cost of this is to make it much more difficult to perform most other kinds of recognitions e.g., to tell if a picture contains two sub-pictures that share such-and-such a relationship. Indeed, holograms may be nearly the worst possible way to represent relations among the features of the things it represents, because it makes it so hard to access those features separately. That's why it is better to represent things as objects–that is, to classify situations into clear, distinct varieties. Otherwise, a memory can't learn: two holograms won't match at all unless the two entire scenes are virtually identical.

Memory and learning are useful only if they represent relations that are partially predictable. They simply can't depend on all the arbitrary features of a situation. If a scene contains 50 features, then there are a quadrillion subsets of features. Without some grouping idea like the concept of object, which makes some subsets predictable, we'd never see the same thing twice, hence could never learn from experience. Then knowledge can't accumulate.

CAUSES and GOALS

How does knowledge help? That question may seem frivolous unless we recognize that no two problems ever are the same in all respects. Thus, past experience can have no relevance–unless we have a way to see some aspects of the world as staying the same, while other aspects change. That's why we can't have knowledge without some ways to see the world in terms of "predictable," descriptive elements–that's what I meant by objects. Furthermore, knowledge can have use only if we can discover suitable couplings between those predictable-features and the Actions we can take. Only then can we learn which actions can make undesirable features disappear. Every evolutionary mind-development must seize opportunities to discover sensory-action correlations that enhance the animal's survivability. The most powerful such discoveries are those which can lead to making predictions based on contemplated mental action-chains–that is, on the ability to make plans.

To say "Y happened because of x" is, in effect, to say that x is a feature with some distinction in regard to predicting which actions can lead to Y. To learn to control the environment, an animal does better by finding better "causes"–fragments of better than chance predictions. But such predictions are computationally infeasible when too many small effects "add up" to cause the changes we perceive–because the number of combinations to keep track of grows exponentially with the number of features. When many features interact, so that we can't see what "causes" things to happen–that's when we call a problem "hard!" Then, so far as I can see, there's just one way to proceed: reduce that complexity by "thinking."

To deal with something complicated, one must find a way to describe it in terms of sub-structures within which the effects of actions tend to be localized.

A problem seems hard when it isn't obvious what to do! The most general way we know to solve problems is to set up a system that has a sense of making "progress toward a goal". In the late 1950's, Allan Newell and Herbert A. Simon worked out a theory of what they called the "General Problem Solver"–a theory of how to reach a goal by "making progress" by finding actions which can replace each problem that has a "high-level difficulty" by other problems which each have lower-level difficulties. I see no way to prove that all intelligent problem-solvers, however alien, must use this selfsame principle. But until we find another idea of comparable power–and none seems on the horizon–this one seems both so simple and so powerful that it is hard to imagine as intelligence evolving without discovering and using it.

RELIABLE COMMUNICATION

Before we ask how aliens communicate, we ought to ask how humans can. Is ever there a word whose meanings are precisely the same for two of us? We each have sometimes wondered, "could two people have quite different meanings for their words, yet never sense that anything was wrong?" What if each thing we both agree is "green" were "really" blue to you and green to me? The Sparseness theory claims we needn't have much fear of that, since, probably, one of those two outwardly indistinguishable meanings will be vastly more complicated than the other. So both of us will almost surely build the simplest one first. Sparseness means we can trust one another.

Of course, we know very little about where this leads, because we still know so little about the details of how sparseness isolates particular concepts. My intuition is that it will indeed support the logical, mathematical and physical arguments Freudenthal proposed in LINCOS–even to include the miniature models he makes for discussing social and administrative matters. But introspection is no guide for guessing which of our common sense concepts are really "simple," because many things we find easy to do use very complex brain-machines that have weak interactions with our "thinking parts". We feel that it is easy to stand on two feet–but some aliens might find this quite astonishing.

What other kinds of ideas might be nearly so universally easily–as islands or markers in that great sea of all conceivable ideas? Surely this must include such concepts as utility, linear approximation, and simple program-like processes, which we could use to communicate about arrangements for trade and commerce. This might also include some basic facts about biology–along with some computational concepts that relate to various aspects of mentality–such as ideas about objects, goals, and memory. But even communicating these might involve all sorts of obstacles. At some point sparseness has to fail, where complicated things can have all kinds of variations and alternatives.

There's little more to say of this today, with any scientific certitude. Tomorrow, with those soon-to-come enormous gains in computational power, we may be able to explore just a little further into the mysterious ocean of all possible simple machines, and perhaps see a few more ideas that are isolated enough to share with other minds. That exploration, too, might tell us more about the origin of life itself, by showing us the simplest schemes that could support first stages of an evolutionary search.

REFERENCES

Freudenthal, Hans. LINCOS, North-Holland

LINCOS drafts a detailed scenario for communicating with aliens. He begins with elementary mathematics and shows how many other ideas, including social ideas, might be based on that foundation. Some of Freudenthal's constructions seem very profound.

Lenat, Douglas. AM and Eurisko. The computer program 'AM' discovered many principles of arithmetic in the course of an evolutionary search.