

If A Computer Speaks In The Forest

On Human Misunderstanding

Tom Sgouros

May 3, 2002

Intelligent About What?

I live in New England, and years ago, I was an enthusiastic fan of the local sports teams: the Red Sox, the Patriots and the Celtics. But the Red Sox broke my heart in 1978, when they blew a 14-game lead to lose the American League pennant to the New York Yankees. Coincidentally, this was a year that I found myself far from home in a community lousy with Yankees fans. The combination was too much, and I haven't followed them closely since. The Celtics began their decay some years ago, and haven't been nearly as much fun to watch as during the 1980's, when I was hooked. And somehow the violence of professional football no longer speaks to me in the same ways that it did when I was fourteen, and so I've lost track of the Patriots, too. (Though the last Super Bowl inevitably caught me up to some degree.)

The net result of this is that there exist bars in which no human present would think me capable of intelligent conversation.

I am intelligent about some things, even if my sister thinks this is a matter for dispute. But I am only intelligent *about* certain things. I am intelligent about things that are part of my own sensory experience, or that I can find some way to relate to that experience. If, by the standards of another person, there are certain holes in my experience, those holes represent subjects about which that person and I cannot talk in a manner satisfying to each other, unless it is in the context of teacher and pupil.

Now assume for a moment an intelligent computer. That is, a computer capable of observing its world, and learning and creativity. What could that computer have to be intelligent *about*? Most computers I know have nothing in the way of senses. If I could say anything about the senses of the one in front of me, I could say that it senses the keys I hit, and it could probably be quite intelligent about patterns of bits. But that's about all. Even I can think of more interesting things to say about professional sports than this computer could ever hope to think

of about any human subject.

Now suppose I give my computer an eye. A camera feeding directly into its innards. This is a start, but it's only a start. In addition to the eye, we have to figure out how to give the computer a richness of experience profound enough that it can start to draw conclusions about what it "sees." Experience and genetics have taught my eye to detect motion and color, but also to correlate those senses with others, and with short-term predictions of the future. I can see wind on water and correlate it with the feeling on my cheek. I can see a glass of beer coming my way and correlate it with either the taste of the beer or the need to duck, depending on how fast it's coming at me. Most important, experience has taught me — and continues to teach me, I am sorry to say — where it is important to pay attention in a situation.

When we count five as the number of senses we have, we are guilty of a vast undercount. At the simplest level, my eyes can detect motion, brightness, color, distance, and pattern (or its absence). They do this by analyzing the light that falls on them, so we call these all aspects of the same sense, but they are truly different senses, as anyone knows who has tried to endow a computer with sight. And this list is not exhaustive. We can debate about whether these are part of my sensory or my cognitive apparatus, but let's point out that my eyes can also recognize faces, read letters from the roman alphabet, and warn me of certain kinds of danger. Some of the "processing" for these stimuli may even happen in my eyeball, before the "signal" reaches my brain.

After we're done pointing out the many ways in which the canonical five senses can be subdivided, let's also note that my immune system is also constantly detecting chemicals as I contact them, and that there is substantial interplay between the neural and the immunological (Gilbert, in press), with who knows what effect; that I may or may not have a vomeronasal organ to detect pheromones; and that the nerves that send me signals from my stomach, lungs, bladder, muscles and intestines aren't necessarily

covered under “touch” in that canonical list.

The point is that there isn’t yet a computer around that can even begin to approach the range of senses I can muster. The range of experiences possible with these senses would be difficult, perhaps impossible, to convey to anyone who can’t experience them himself. I don’t know how I would ever describe them to a computer. What language could I use?

On the other hand, I’m not sure that I have any way to interpret what may be a perfectly intelligent statement about a series of bits, or to appreciate the varieties of pain caused by a corrupted disk file.

This sounds flip, but consider it anyway. A computer might be perfectly intelligent — able to make inferences about what it experiences, and to think creatively about the future — but how would I know? How would it express these ideas to me? How could it even begin to describe to me things I can’t feel?¹ Without a shared experience on which to ground our conversation, significant communication cannot happen. My children and I can only really talk about things that we share, or can imagine sharing. It’s hard to imagine what experiences I could share with an eyeless machine in a stationary steel box.



A great deal has been written about natural language processing in computers and about how to get a computer to handle a natural language in a way that could fool a human into thinking she was talking to another human. The prospect of inventing a machine to understand spoken commands or to translate English into Finnish is so appealing that it has supported a forty-year attempt to seek the mechanical underpinnings of the language we use to talk to one another. Sadly, after forty years of trying, no one has yet succeeded. We have machines that can transcribe spoken speech, and match it up with pre-existing command syllables, and we have lexicon-translators that can do a fair job translating simple sentences from one language to another, but there is no machine yet that can manipulate day to day language even as well as an average three year old.

Fifty years of molecular biology have brought us to the point where it is possible to see where the lure of abstract information have and have not been helpful to progress in the field. Despite the almost perfect analogy between

¹For that matter, how does it even know I am here to hear them? I tap at its keys, but why not interpret that as friendly patting, or some kind of environmental white noise? It took Helen Keller quite some time to realize that Anne Sullivan was trying to communicate with her by patting her hand. How long will it take a machine?

DNA base sequences and digital data the puzzles of heredity have proven to be only partially responsive to solutions proposed by applications of information theory. The conclusion to which biology has been forced is that it may not be tremendously useful to think of DNA as “information,” and there may even be good reasons to avoid the characterization.²

Perhaps it is time to conclude something similar about the structure of the language we use: that while there are surface structures strongly suggestive of a formal deep structure, strongly suggesting something is not the same as being it. That is, there is a good analogy to be made between natural and formal languages, and while it is a good analogy, it is only an analogy, and cannot be pressed too far. Let’s look briefly at the original source of the idea.

Information, Please

In his classic paper, which became the foundation of the field of information theory, Claude Shannon wrote this:

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. (Shannon, 1948) [emphasis in original]

We are talking then, about messages whose content comes from some set of possible messages, and whose form is entirely independent of content. That is, whose “syntax” is unrelated to its “semantics.” Seen as a delimitation of the class for which the theory applies, this is unexceptionable. It’s the beginning of the presentation. But it’s hard to square the words used with that limited intent. Shannon seems to speak with much broader intent, to cover all aspects of communication between two “entities,” be they animal, vegetable, or mineral, and that seems to be the way the theory was widely understood.

But these are Shannon’s assumptions, not his findings. Does communication really work this way?

²See, for example, Lewontin, 2000, or Venter *et al.*, 2001, p.1346.

In many cases, the answer is no. Look for a minute at the simple case of one computer transmitting some data to another. Let's hook them together by a simple serial (though by now slightly archaic) RS-232 link. One might consider the "syntax" of messages transmitted along this line to be that they are composed of eight-bit bytes, and each byte is transmitted as an oscillating voltage, according to rules governing timing and voltages.

It is common, when transmitting large chunks of data, to put them after a description of the data to follow. For example, to send 100 16-bit integers in a way that will keep it from being interpreted as a sequence of 50 32-bit integers or a string of 200 8-bit characters, a computer might put something like this in the front of the message:

```
Int 16 [100]
```

The character string of text the same length might be indicated like this:³

```
String [200]
```

These indicators (sometimes called "metadata") are embedded among the data transmitted, but are meant as an indication of the kind of data that make up the actual message. Are they then a part of the message, describing another part, or are they an essential part of transmitting integer or character data? Are they part of the meaning of the message or the syntax of the transmission protocol? Some say one and some say the other. It seems mostly to depend on the level of analysis you choose.

The potential complexity of this simple example is dwarfed by technologies in daily use by millions. Data transmitted across the internet is divided and encoded into "packets," which are taken apart and reassembled according to instructions transmitted alongside the data itself. These packets are themselves used to enclose encoded data, such as "MIME documents," which contain their own instructions for unpacking. The overall picture is quite intricate; internet design documents define six distinct layers of abstraction above the basic level of oscillating voltages connecting two computers.⁴

³Of course this isn't international. Ideographic languages, for example, need many more than 8 bits of information to specify a single character.

⁴In case you're wondering, they are: the physical layer (kind of cable connecting neighboring machines), the data link layer (the kind of communication between one machine and its neighbor), the network layer (how the computer network is laid out), the transport layer (how does data get from one computer to another), the session layer (how does a program on one computer cooperate with a program running on another), the presentation layer (what kind of data is being received), and the application layer (how should that data be displayed to the user). You can find more information in IEC, 1994. As with the other domains under discussion here, the standard outlined by this scheme is often honored in the breach.

In addition to all these levels of abstraction, the data you send can be pretty baroque, as well. Common data formats, such as internet graphics formats like GIF or JPEG, or document formats like Postscript or PDF files are filled with this kind of metadata: descriptions of the data enclosed in the message. From one point of view, all the bytes in a Postscript file are data, but from another, only the ones that represent characters to be printed are data. So which is it?

An important point here is that this uncertainty of interpretation has nothing to do with humans. The boundary between syntax and semantics varies according to whether the machine interpreting the message is a line amplifier, an Ethernet router, a firewall, an email reader, or a workstation displaying some document to a human. Each of these machines has a different opinion about what constitutes a message and what constitutes the syntax of that message. The result is that the levels are fairly confused, and what is perceived as the meaning of the message at one level is quite often regarded as part of the syntax of a higher level message.



So if we can't differentiate between syntax and semantics even in machine-to-machine communication, what hope is there when a human is involved? People have been working for years trying to uncover a generative grammar for English — a systematic definition of the syntax of the language, independent of the meaning of its vocabulary — without ever quite being able to define away places where the meaning of the words involved clearly influences the structure of the language. To cite just one well-known example, from Terry Winograd (Winograd, 1972):

The city council denied the women a permit because they feared violence.

The city council denied the women a permit because they advocated violence.

Here, the grammatical role of "they" — does it refer to the women or the city council? — can only be resolved by someone who knows not only the meanings of the words involved, but also something of city councils. There are no clues in the the formal elements of the sentences; you need to know about the world, too. You cannot resolve the syntax of these sentences without experience of the world.

This is not a problem limited to esoteric example sentences dredged up from the minds of linguists with nothing better to do. These issues crop up in speech most

people encounter every single day. Slang meanings for common words abound, for example, and context is the key (sometimes the only) clue to which meaning is meant for a word. Homonyms, ambiguity, and how far to extend universals (does “everyone does it” *really* mean that everyone does it?), also require a sensitivity to context to be understood correctly.

Finite Language?

For natural language, the second of Shannon’s suggested restrictions is also a problem. He restricts consideration to messages selected from some (not necessarily finite) set of possible messages. He talks of language as being a “discrete sequence of symbols chosen from a finite set,” and further makes clear that he regards there to be no essential difference between considering the symbols of language to be letters or words.⁵ But this is a vast oversimplification of how language works.

For one thing, it is an error to imagine that words are usefully comparable to formal symbols. This statement will incline some to reach for a dictionary, but let me forestall that attack by consulting some dictionaries in support instead. A few dictionaries sit quietly on my shelf, but if they could speak, they’d argue with each other. I have one dictionary (*Webster’s New Collegiate*), for example, that counts eight intransitive and seven transitive meanings for the word “walk,” including the baseball meanings. Right next to it on the shelf sits another (*The Oxford Universal*) that counts 20 intransitive and eleven transitive meanings, including “to cheat at cribbage by moving your opponent’s pegs,” but not mentioning baseball. If “walk” can be construed as a symbol, it is one of some fluidity.

The people who write dictionaries know about this, and dictionary writers have been known to cause controversy, at least within the teapot where tempests of that sort rage. *Webster’s Third New International Dictionary* was (is?) controversial when it came out in 1961 because it appeared to ratify in an official organ many words that had been construed as poor usage in the past.⁶ This kind of controversy would be impossible if there was consensus over what words mean.

Polysemy is another kind of ambiguity that makes it

⁵Shannon, 1948, p.5ff. It doesn’t matter that much for this critique, either. Given all the different sounds it can indicate, what does a “P” stand for?

⁶David Foster Wallace has a good account of some ongoing controversies of this sort. (Wallace, 2001) He writes “You can think of *Webster’s Third* as sort of the Fort Sumter of the contemporary Usage Wars.”

even harder to regard words as symbols. Even within the context of an agreed definition, the meaning of a word can vary tremendously depending on its context. For example, many of the qualities of a “conversation” with someone depend very much on who that someone is. There is a difference *in kind* between an adult’s conversation with a child, a friend, or a soon-to-be-former lover, among many others, and these are all subsumed under the same dictionary definition. The philosopher Julius Moravcsik wrote that this kind of flexibility of meaning is an essential part of a language being able to deal with what he called the “constant barrage of small semantic emergencies” encountered each day by people trying make themselves understood. (Moravcsik, 1998, p.37)

The lexicon cannot bear the weight that Shannon loads upon it, and neither can the idea that a sentence is just a list of words. The words in a sentence are crucial, of course, but so are the relationships between those words, and between those words and the context in which they are spoken. Ignoring these relationships is the classic error of reductionism. The atoms of a sentence are the easy thing to see and count, but they are far from the only story. The base sequence of a strand of DNA is easier to see and quantify than the array of enzymes and cellular structures that surround it, but that doesn’t mean that it’s the only important part of the story of development.

To look at another example from the physical sciences, the important thing about analyzing the behavior of a gas is the relations between the molecules that make up that gas, and less so the individual molecules themselves. Similarly, the relations between words — layers of metaphor, allegory, metonymy, catachresis, hyperbole, and the rest of the rhetorical menagerie — create a large part of the meaning in everyday speech. The names seem like dusty relics of high-school literature classes, but we all use them every day: “The White House is in a tizzy over all the press leaks.” These are not just flourishes. Figures of speech can create entirely new meanings,⁷ and it is not at all obvious that all the possible meanings can be adequately described as a “set” in the way Shannon means it.⁸

Time for an illustrative digression: once upon a time, my household was entertaining a visitor of one of my roommates. We were all making dinner together, and hav-

⁷See for example, Kittay, 1987 on metaphor.

⁸Incidentally, it is important to note here that Shannon may not even bear the lion’s share of the responsibility for what was made of his ideas. Others, such as Norbert Wiener, John von Neumann, Walter Pitts, and their intellectual heirs, bear as much, or more. But though others elaborated the concepts beyond all reason, much of the program was spelled out in the original paper. For the sake of this essay then, consider “Shannon” to be a species of synecdoche.

ing a good time teasing one another. This was the normal mode of discourse in our house, and we thought nothing of it, but apparently our guest was disconcerted, and she, a great proponent of new-age counseling styles, said at one point: “Tom, I want you to do me a favor. The next time you think of something teasing to say to me, I want you to make it a compliment instead.”

Well, we were all quiet for a little bit after that, but minutes later she spilled water on her lap when she went to inspect the bottom of the mug she was holding, and I tried to hold back, but it was no use, so I did what had been asked, and said, “Arlene, that’s a really nice sweater you’re wearing.” A moment later some errant spaghetti splashed sauce on my glasses, and she very politely took the opportunity to compliment my haircut. So for the rest of the evening, not an insult could be heard in our kitchen, but the compliments — and the laughter — flew unusually thick and fast.

In what way could the meaning of these sentences be formally related to their structure? How could these meanings possibly be analyzed independently of their context? My sentence wasn’t even untrue; it *was* a nice sweater. In fact, as the rules of this game evolved over the evening, the compliments bestowed had to be true (more or less). The unintended effect of the whole situation was to grant an entirely new meaning to a whole class of sentences — compliments — and the ramifications percolated through our household conversation for the next few days, whenever one of us paid a compliment to another.

Because the meanings of sentences depend so crucially on the world in which they are to be interpreted, there is no way to enumerate their meanings. The set of all meanings is not just infinite, like the integers, but uncountably infinite, like the irrational numbers. There is therefore no way to fit them into the kind of set that would satisfy the limitations of information theory.

This is a good thing, too, since it’s hard to imagine how civilization could have come to be if all the possible meanings one stone-age denizen might have wanted to convey to a companion could have been enumerated in the language the two of them used on one another. Language and civilization grow and change as people think of new meanings to convey to one another.

What’s In A Name?

The causal theory of names has it right, but only for half of the story of meaning.⁹ Following this theory, “Socrates” refers to the philosopher because his parents said so, and they told him and he told everyone else and then someone told Plato, and on and on. Presumably something similar happened to other words, and so we have our lexicon.

The catch, of course, is that the causal chain is not a row of dominoes falling over. Each link in the chain is a person, with his or her own hopes, dreams, prejudices, agendas and even speech impediments, connected to the next person in the chain by a generally imperfect method of communicating ideas. For a chain above a certain length, it seems improbable to imagine that perfect communication could take place. Historians know this, which is why the best histories rely on multiple sources. Children know this; they play “telephone,” whispering a phrase into each others’ ear around a large circle to see how the message is mangled.

Consider the word “horse.” Presumably there was a first somebody who gave a name to those big things lumbering around the steppes. But the name probably didn’t sound much like “horse.” The causal chain from that person has involved quite a bit of evolution of the word, changing its sound, and defending territory from the Latin “*equus*” and the Greek “*hippus*,” and what’s more, changing the category of animal implied by the word. The horses that we think of today are not very much like the horses of the steppes. Further, we now differentiate between horses, donkeys, burros and ponies.

But what is a horse to *me*? Aside from a terrifying ride on one arranged by my third-grade teacher, I had no contact with horses when I was growing up. My concept of a horse is mostly from what I read, and in quality not much different than my conception of an aye-aye. But someone who grew up riding horses would have a much more visceral definition of one. That person might have an intricate web of sensory memories about horses, where all I have is images from movie Westerns, evolutionary charts, the theme music from Mr. Ed, and a dictionary definition.

The evolution of the world and the reality of linguistic evolution dictate that it cannot be the case that a word has a timeless meaning, but that it is given its meaning by association with the world experienced by the person who wields it. That world contains both sensory experience and the experience of other people’s words. Some words are defined mostly by experience: nose. Some are defined

⁹See Kripke, 1980 and Putnam, 1988 for the original elaborations of this theory.

mostly by other words: justice. And most are defined by a mixture of the two.

The fact that words are a mixture of the two is an important point. There is a risk of infinite regress to imagine that words are only defined in terms of other words. But this is not true if we look at different words as containing different proportions of “experience-meaning” and “dictionary-meaning.” As Dennett pointed out in relation to the concept of the homunculus, there is no infinite regress in imagining nested homunculi, so long as they get progressively dumber as you go down. Similarly, so long as the proportion of experience-meaning grows as you descend the levels of dependent meaning, you’re fine.

An inescapable ramification of considering experience-meaning to be at the heart of our language is that the meanings of my words differ from the meanings of yours.



The fact that we can still use the same word to cover our very different meanings says more about the social conventions of language than about shared meaning: it is relatively easy to satisfy each other that we are talking about the same thing, even when our internal definitions differ so profoundly.

As to the problem of how we communicate when meaning is so variable, well there are rules of language we can follow when we want to be understood, but mostly it’s a hit and miss kind of affair. Let us each consider our own experience and personally refute the misconception that the vast majority of human interactions are characterized by a perfect understanding between the parties. The only people who could possibly really believe this are people who have never been on a date, who have never had children, who have never really considered what goes on in order to condense real events into stories in a newspaper.

We fool ourselves into thinking of language as an amazing tool of communication by constructing and analyzing sample sentences about hills and balls, and imagining personal interactions as shallow as a transaction at a convenience store. But what do people really talk about? Food, love, politics, religion, other people, the stock market, the economy, the weather. In any of these realms of discourse, misunderstandings are the rule, not the exception. The amazing part is that, given the varied experiences we bring to most conversations, communication can happen at all.

A Logical Digression

A “formal” logical system is one in which one can effectively distinguish between the meaning and structure of assertions. A system of this sort consists of two parts: the rules you can use to manipulate and combine assertions in that system to create other assertions, and the axioms, which are the assertions at the base of the system from which all true statements in that system are derived.

In trying to posit a set of rules and axioms for a particular observed system, there are several traps for the unwary. One is that the axioms and rules useful for one scale of analysis may not be applicable to another. We saw this earlier in the examples about machine-to-machine communication. For any given piece of internet hardware, there is a set of applicable axioms and rules with which one can generate possible messages in that system. It is, however, a mistake to try to use axioms useful to talk about the transmission of bits to try to analyze the structure of a Postscript file.

Another trap involves the number and character of the axioms themselves. To resolve the above example about city councils and violence, we could add an axiom to our system about whether city councils are ever likely to advocate violence. The problem then becomes setting an upper bound on the number of axioms. Given the variety of possible experience, it is highly unlikely that we could, even in principle, do so.

When I listen to someone else (whose speech I understand), I am, in some sense, processing the meanings of the words, as well as the additional meaning imbued by the structure of the sentences I hear (as well as meaning from the intonation, the situation and so on). But I am also capable of inferring new grammatical axioms to use in my understanding. When I first heard someone put “not” at the end of a sentence, I was puzzled only for a moment before I figured out what was being said, and why, and I suspect the same is true of most people with an adequate command of English. You hear a new rule, understand it, and (if you decide you like it) you use it again later.

The process of reading *Riddley Walker* (Hoban, c1980, 1998 printing) involves, in the early chapters, constant inference about the usage of the language used. It’s close enough to English that you can guess what’s up, but far enough that it’s clearly a stretch.

Any particular brain is finite, so presumably can only accommodate a finite number of axioms. But it is nowhere written that the axioms in my head are the same as the ones in yours. If we posit, for the sake of argument, the existence of a generative grammar for *my* language — a set of axioms and rules of manipulation that can be used

to generate all the things I can think of saying — it would not be the same structure as the one that controls *your* language. If there is such a thing as a generative grammar, it must differ from one person to the next.

If my generative grammar depends on my understanding of English words and the world I live in, then it can't be the same as yours. Further, I can't have been born with it intact. Similarly, though one can find many parallels between languages, it isn't possible to claim that at root they're all the same.



When Gottlob Frege wrote about the possibilities of a formal logical system, he only considered systems with a finite number of axioms. Later, the work of Hilbert, Gödel, and Turing extended this to an enumerably infinite number of axioms.¹⁰ But it cannot be the case that a formal system — one where you can plausibly separate syntax and semantics — contains an uncountably infinite number of axioms.

So perhaps this is both the answer to why it appears that a generative grammar is possible, and why it seems so hard to actually achieve. The rules of grammar are finite in any given individual, which implies that the separation between syntax and semantics is perfectly plausible (though still debatable). However, so long as people remain creative, and continue to be born, the number of rules is, in principle, infinite across the society of common language speakers, which is where the assumption breaks down.

In the context of Frege and systems of formal logic, “finite” means finite in principle, whereas in the context of cognitive psychology, “finite” means only finite in this particular brain, or at this particular time. To claim that the English language, for example, is bounded by a finite set of axioms is to deny the obvious reality of the evolution of that language.

We can compare Chomsky to Thor visiting Utgard. He was challenged to empty the giant Skrymir's drinking horn in one gulp, but didn't realize it was connected to the sea. Even so, he drank so much that the ocean ebbed. Chomsky, similarly, has spent decades trying to codify the apparently-finite, but in-principle-infinite. The work has profoundly deepened our understanding of language,

¹⁰A formal system can accommodate an infinite number of axioms only if they are generated from a finite number of axiom schema. So, for example, you can have an infinite number of axioms of the form “six is a positive integer” and “seven is a positive integer,” and so on. Turing stated the requirement in terms that are more familiar now: a formal system is one where axioms can be — at least theoretically — combined by a computer to create proofs.

but the original goal of providing a systematic account of the deep structure of all human language no longer seems reachable.

But let's reject Manicheanism. Just because a generative grammar may not be possible in the way it's been proposed doesn't mean that elements of one are not present in our brains. The idea wouldn't have made such an impact if it hadn't had tremendous explanatory power when applied to real data. This is not the place for a serious examination of alternatives, but maybe instead of an innate grammar mechanism, there's an innate grammar-inference mechanism, that children use to deduce grammar rules. Some do it better than others, but then we already knew that: there must be some reason that some people speak or write so much better than others.

To imagine people communicating who don't really share the core meanings of their language, but are the products of parallel development seems unwieldy and odd, but so does a moose, and yet there they are, all over Maine. As we've learned time and again, natural systems are not bound by our conceptions of order. Besides, it's not all bad: Harold Bloom wrote that misunderstandings form the basis for some of our best art.

There's a comparable thought in developmental biology. Biologists frequently talk about the cells of a growing organism being “commanded” to do this and that, but it's hard to figure out where the commands come from. Perhaps it's better to say that a cell in this or that particular situation “decides” on a particular course of action.¹¹ It seems hard to imagine billions of cells independently deciding things and making a person out of those uncoordinated decisions, but again, there is the moose to consider.



So I'm back to imagining an intelligent computer, able to wander freely about the world, experiencing what it can, learning what it can, and thinking only what it can. Were I to meet that computer in an apple orchard, and watch it select a ripe apple from a tree, and pick it, and appreciate the redness of it and the fine fall weather we're having, I would still be pretty uncertain that it could appreciate the apple in any way remotely similar to the way I experience an apple.¹² And I think it would probably feel the same way about me.

¹¹See, for example Black, 1998.

¹²This is not to mention my capacity to *enjoy* the apple. Damasio, 1999 argues persuasively that in some ways, emotions may be even more fundamental to consciousness than our senses. See Leiber, 1996 for more evidence to support the view that senses may be secondary.

My own opinion is that anyone who makes confident predictions about the inability of engineers and scientists to achieve some specific goal is sticking their neck out into places I wouldn't dare. The lessons of the history of engineering, and particularly computers, seem clear to me: naysayers eventually get buried. Didn't I read that the chairman of IBM once predicted a worldwide market of a handful of computers?

So I will then assume that someday there will be built an intelligent computer, capable of inference, learning, and creativity. Here's my prediction: when we get one, learning to communicate with it in any way that would suggest that it is, in fact, intelligent will be at least as challenging as building it in the first place. When at last we achieve a machine that can use language to express its own feelings and experiences, we won't have the slightest idea what it's talking about.

References

- Black, Douglas L. 1998. Splicing in the inner ear: a familiar tune, but what are the instruments? *Neuron*, **20**(February), 165–168.
- Damasio, Antonio. 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Harcourt Brace.
- Gilbert, Scott F. in press. The role of predator-induced polyphenism in the evolution of consciousness: A Baldwinian speculation. In: Weber, B. (ed), *Evolution and Learning: The Baldwin Effect Reconsidered*. MIT Press.
- Hoban, Russell. c1980, 1998 printing. *Riddley Walker*. 2d (expanded) edn. Indiana University Press.
- IEC. 1994. *Information technology – Open Systems Interconnection – Basic Reference Model: The Basic Model*. Tech. rept. International Electrotechnical Commission.
- Kittay, Eva F. 1987. *Metaphor: Its Cognitive Force and Linguistic Structure*. Oxford University Press.
- Kripke, Saul A. 1980. *Naming and necessity*. Harvard University Press.
- Leiber, Justin. 1996. Helen Keller as cognitive scientist. *Philosophical Psychology*, **9**(4), 419–440.
- Lewontin, Richard. 2000. *It Ain't Necessarily So: The Dream Of The Human Genome And Other Illusions*. New York Review Books.
- Moravcsik, Julius M. 1998. *Meaning, Creativity, and the Partial Inscrutability of the Human Mind*. CSLI Publications.
- Putnam, Hilary. 1988. *Representation and Reality*. MIT Press.
- Shannon, C. E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal*, **27**(July, October), 379–423, 623–656.
- Venter, J. Craig, *et al.* 2001. The sequence of the human genome. *Science*, **291**(16 February), 1304–1351.
- Wallace, David Foster. 2001. Tense Present: Democracy, English, and the Wars over Usage. *Harper's Magazine*, April.
- Winograd, T. 1972. *Understanding Natural Language*. Academic Press.