Framed: Language in the Artificial Intelligence Debate

by

Dr Ron Gallagher
P.O. Box 118
Elwood, VIC 3184
Australia

CONTENTS

Introduction

Artificial intelligence(AI) robots have difficulty keeping track of a changing environment: this is the frame problem. It is proving difficult to equip AI robots with models of their individual worlds and the capacity to navigate everyday situations. Considered more broadly, the frame problem poses important questions about how human beings acquire knowledge of the world, and use common sense to deal with the unexpected.

Daniel Dennett describes the frame problem as a “new, deep epistemological problem - accessible in principle but unnoticed by generations of philosophers - brought to light by the novel methods of AI, and still far from being solved.[1]

Science fiction writers began to address this problem as soon as they imagined the first robot. For decades, science fiction has used robots and computers to explore what it means to think like a machine and what it is to be a human being.[2] Science fiction dramatises philosophical problems in a way which makes them accessible and concrete. By clothing these issues in the practical and ethical concerns of human (albeit fictional) protagonists, science fiction brings philosophy alive. A secondary effect of setting up these *philosophical thought experiments* is that it often makes the problem dissolve before our eyes. Dennett notes that many of the thought experiments which philosophers are fond of using to illustrate their arguments are flawed because they are not fully visualised. There are gaps which the reader is supposed to fill in with intuition and imagination. Examples of these famous thought experiments are Searle’s Chinese room and the Brain in the Vat problem.

Epistemology is concerned with how we can know things. It asks "What is knowledge?" and "Is knowledge possible?" In particular epistemology seeks to establish grounds for knowledge. Dennett's concern about what constitutes a "fact" highlights uncertainty as to what may be considered knowledge.

It is generally accepted that knowing "how" is different from knowing "that"; the former is concerned with ability, the latter with truth. Dennett's objection suggests that knowledge only arises when something is being done. Knowledge cannot be considered as an abstract entity, it is always a function of its context and application.

[1] Daniel Dennett. “Cognitive wheels: the frame problem of AI", *Minds, Machines and Evolution,* edited by Christopher Hookway, Cambridge University Press, 1984. p.130.

In *Consciousness Explained,* Dennett examines how the gaps in these scenarios misdirect the reader in much the way a magician misdirects an audience when performing a trick. Accusing the philosophical fraternity of sleight of hand is a fairly serious charge, but if science fiction in any way provides a remedy for such philosophical misdirection - by fully visualising the scenario - then it must be welcomed as a valuable philosophical tool. Some of Dennett's critics have accused him of using science fictional examples to patch over gaps in his argument. I believe the opposite is the case - the science fiction examples clarify the argument. Science fiction explores many 20th century themes in depth and breadth. It provides powerful models for the exploration of scientific, anthropological, social and political themes. In fact, science fiction uses the most ancient of narrative structures to complement some of the most incisive thinking of the 20th century. Science fictional themes are also the themes of Darwin, Einstein, Martin Luther King, Richard Dawkins, Stephen Hawking, and other great 20th century thinkers. Science fiction deals with ecology, genetic engineering, overpopulation, war, and most of the other pressing issues of our century - and adds a crucial human element to the exploration of these issues.

John Searle's Chinese Room
Searle imagines a person locked in a room with a series of Chinese symbols in various baskets, and a rule book, written in English, explaining how to manipulate these symbols. When Chinese symbols are passed into the room, the rule book indicates which ones are to be passed out. The person in the room does not know that the symbols passed in are questions, and the symbols passed out answers. The programmers are so good at designing the program that the answers are indistinguishable from those of a native Chinese speaker. He writes, "The point of this whole story is simply this: by virtue of implementing a formal computer program, you behave exactly as if you understand Chinese, but all the same you don't understand a word of Chinese." As far as Searle is concerned, an artificial intelligence simulates understanding in the manner of the Chinese Room.

[2] The short story anthology *Machines That Think* is an excellent survey of the depth and breadth of these explorations. *Machines That Think* edited by Isaac Asimov, Patricia S.Warwick, and Martin H.Greenberg. Harmondsworth: Penguin, 1983.

Over the last 10 years I have used science fiction texts in my teaching, to explore all aspects of twentieth century thought. I confine myself in this book to exploring how the frame problem highlights issues in linguistics and the philosophy of language, and use science fiction to illustrate and clarify, solve, and even dissolve, philosophical problems.

The Brain in the Vat
Suppose evil scientists removed your brain from your body while you slept, and set it up in a life-support system in a vat. Suppose they then set out to trick you into believing that you were not just a brain in a vat, but still up and about, engaging in a normal embodied round of activities in the real world. How would you know? Might you be now, and have always been, a brain in a vat?

The most important clarification I would like to make at the outset is that language is not primarily a means of communication. Many animals communicate without a fully-fledged language. The most decisive feature of language is not its role in communication, but the fact that it enables one to talk to oneself. By talking about what we are doing and what we are going to do, we plan our engagements with the world. This ability to represent the future to ourselves is generally agreed to be the crucial factor in intelligence and consciousness. The fact that it is language which enables us to do this, elevates language in debates about intelligence and consciousness. It has often been argued that if computers and chimpanzees could use natural language - English or Spanish, for example - they would be considered intelligent. The criteria of intelligence in these cases has always been the agent's ability to communicate. Several computers and primates have demonstrated the ability to converse - but have not been considered intelligent.[3] The goalposts have been moved yet again in the intelligence debate and now it is an agent's ability to plan for the future which is considered the crucial criteria of intelligence. The frame problem is precisely the problem that artificial intelligence robots and computers have when attempting to plan their next move.

If an AI robot had natural language, like that of human beings, it could plan ahead and be considered intelligent. But enabling a robot to use natural

[3] See Weizenbaum(1984) for a discussion of ELIZA a computer programme which emulates a therapist. Also see Savage-Rumbaugh(1994) for a discussion of language learning in primates.

language is proving very tricky. Just getting a computer to understand colloquial English is an elusive goal. Even if a computer could understand English there is the additional problem of enabling it to speak to itself and plan ahead. No one has yet equipped a robot or a computer with this ability to analyse its own position in the world and act on it. The majority view in AI is that the robot needs a model of its world before it can act in that world successfully. Robots which merely avoid objects when moving around are not considered intelligent. A robot which organises its world or its actions so that it does not have to continually avoid objects, might be considered to be on the way to intelligence. This simple goal of AI has not yet been achieved.

Dennett describes our ability to talk to ourselves, and thereby plan for the future, as a "good trick". He doesn't reveal the secret of this trick, but there is a great deal of speculation about how it works. One theory is that language[4] provides a kind of model of the world, or has a relationship to the world comparable to that of map to ground. There is also a strong view that the grammar of a language can define relationships in the world, and that these grammatical structures map onto the world in ways that reflect relationships in the world. In this book I argue three main points:-

1. Language does <u>not</u> embody a model of the world.

2. Language does <u>not</u> delimit how or what we can know about the world.

and drawing on the work of Wittgenstein[5]...

3. Language is possible because of agreement about the frame.

The larger AI debate is about the nature of consciousness and whether a robot could ever be considered conscious and truly intelligent. I am happy to

[4] When I use the word "language" I mean natural language in general - not a specific language.

[5] See Ludwig Wittgenstein. *Philosophical Investigations*. Oxford: Basil Blackwell, reprinted 1967. "If language is to be a means of communication there must be agreement not only in definitions but also in judgments." Para. 242.

imagine that a fully functional robot such as Asimov's Andrew Martin[6] or *Star Trek*'s Data[7] could be built - I also agree with the view, which these two robots epitomise, that the issue is not primarily philosophical or technical but ethical. The real issue in the artificial intelligence debate is - could human beings ever accept a machine as an equal, with rights as well as duties? I do not expect to dissolve the artificial intelligence debate in this short piece, merely to clarify the role of language in the debate. I will argue the above three points by extending a thought experiment of Dennett's involving an AI robot and a midnight snack.

[6] "The Bicentennial Man" Reprinted in *Machines That Think* edited by Isaac Asimov, Patricia S.Warwick, and Martin H.Greenberg, Penguin, 1983

[7] See "The Measure of a Man" *Star Trek: The Next Generation* written by Melinda M.Snodgrass and directed by Robert Sheerer, Paramount Pictures 1989.

1. The Frame Problem

In the film *Short Circuit,* the heroine, Stephanie, mistakes defence robot Number 5 for an alien. Stephanie is accustomed to taking in stray animals and is happy to comply with her space visitor's requests for "more input." She shows it around her home, pointing out general features of the house and naming her animals. Number 5 then becomes obsessed with the television, one of our biggest sources of "input", and quickly assimilates the language and expression of hundreds of film characters. It uses these assimilations as "role models" and eventually develops its own theories about life, death and morality.

Despite the vast knowledge acquired from encyclopedias and television over a few days, Number 5's attempts to cook breakfast for Stephanie are fraught with problems. The recipe for pancakes specifies that one should mix milk, flour, and eggs, but it doesn't state that one needs to crack the eggs into the bowl and discard the shells. Number 5's attempts to cook hash-browns are similarly fraught. The instructions read,

> For crisp yet moist potatoes brown on one side then turn over.[8]

[8] *Short Circuit* directed by John Badham, 1986.

The instructions don't mention that you have to remove them from the packet before you grill them! This sticky, smoke-filled scene is a dramatisation of one aspect of the frame problem - Number 5 does not know the difference between the container and the thing contained.[9] Daniel Dennett in his article “Cognitive Wheels: the Frame Problem of AI”[10] provides an example which shares many features with the above scene. In order to highlight aspects of the frame problem, he imagines the problems presented in the making of a midnight snack. Clearly there are some very obvious things one needs to know in order to make a midnight snack; where the fridge is, where the knives, plates and glasses are, and if there is any bread. If, however, the snack was being prepared by Number 5, its programmer might also have to program it with some other data. Dennett comments,

> For instance, one trivial thing I have to know is that when the beer gets into the glass it is no longer in the bottle, and that if I am holding the mayonnaise jar in my left hand I cannot also be spreading the mayonnaise with the knife in my left hand.[11]

Dennett speculates that such knowledge might be innate in human beings,

> Perhaps these are straightforward implications - instantiations - of some fundamental things that I was in effect *born knowing* such as, perhaps, that if something is in one location it isn't also in another, different location; or the fact that two things can’t be in the same place at the same time; or the fact that situations change as a result of actions. It is hard to imagine just how one could learn these facts from experience. [12]

Perhaps human beings do learn these things from experience, but how does one teach a robot to learn from experience? If this were possible, it would be a very long process, something like bringing up a child. One AI solution to shortcut the process is to fill the robot with "input", in the form of facts about the environment that the "intelligent robot" is likely to encounter. Dennett comments,

> I listed a few of the many humdrum facts one needs to know to solve the snack problem, but I didn't mean to suggest that those facts are stored in me - or in any agent - piecemeal, in the form of a long list of sentences.[13]

---

[9] One might also wonder how Number 5 knows that “brown” is a verb, or can understand how something can be crisp and moist at the same time.
[10] Daniel Dennett. “Cognitive wheels: the frame problem of AI", *Minds, Machines and Evolution,* edited by Christopher Hookway, Cambridge University Press, 1984.
[11] Ibid.
[12] Ibid.
[13] Ibid.

Dennett lists a few things that we know.

> We know that mayonnaise doesn't dissolve knives on contact, that a slice of bread is smaller than Mount Everest, that opening a refrigerator doesn't cause a nuclear holocaust in the kitchen. [14]

These facts may seem obvious - common sense - to human beings, but are they really facts? Consider the sentence “When a thing is in one place, it is not in another place.” Does this state a fact? Whether you think it does or not, AI researchers are seeking ways to embody such assumptions about the world in their robotic agents. Dennett concludes that if AI researchers can’t generate all this "common sense" or "frame" information from a small number of axioms, they must devise ways of feeding this vast amount of information to their robot, storing it, and enabling their robot to access it quickly.

Artificial intelligence problems are usually played out in a simplified environment where the robot or computer is set well defined tasks within a well defined environment. The robot has a reduced set of aspects to exercise its reduced set of axioms on. This method of investigating intelligence is, in Dennett's view, flawed because *a fact is only a fact when it is a relevant fact.* Whether one develops a robot that acquires genuine experience or brings one’s robot quickly up to speed with vast amounts of experiential data, the crucial problem remains of how to enable the robot to select the relevant “experience” to apply to a novel situation. Deciding what are relevant factors in a situation is an intelligent act. If AI researchers indicate which information is relevant, they are doing most of the robot's thinking for it. Dennett compares our intelligent robot to a "walking encyclopedia".

> A walking encyclopedia may walk over a cliff, for all its knowledge of cliffs and the effects of gravity, unless it is designed in such a fashion that it can find the right bits of knowledge at the right times, so it can plan its engagements with the real world.[15]

Let us imagine what we need to do in order to build ourselves a midnight snack robot, which we will call Midnight. We probably need to program it with the necessary information to deal with the eventuality of an empty mayonnaise jar, but we wouldn't ordinarily need to program it with information about cliffs and

[14] Ibid. It is easy to place a context around these "facts"; if we suspected the mayonnaise was acidic; if we were in the land of the giants; if we were in a booby trapped kitchen, all of these non, or pseudo, facts become very relevant facts.

avoiding falling over them unless it was about to go on a midnight picnic. Even if we anticipate the possibility of Midnight going near a cliff, or encountering an empty mayonnaise jar, do we need to program it with other "negative facts" by assuring it, for example, that turkeys don't explode when removed from the fridge? The number of such negative facts is infinite - bounded only by the imagination. Such knowledge only becomes relevant if the programmer anticipates a cliff, a poorly stocked larder, or a turkey stuffed with nitro-glycerine.

Cog & CYC: Two Approaches to Artificial Intelligence

There are two main schools of thought in AI research:-

1. the top-down approach - characterised by the work of Douglas Lenat who is filling his AI (called CYC) with encyclopedias of facts.
2. the bottom-up approach - exemplified by the work of Rodney Brooks who lets his robot (called Cog) find out about the world by trial and error.

The former approach is really an extension of expert system development. However, in the case of CYC, the machine is being taught to be an expert about common sense.

The latter approach has much in common with child-rearing. After many years of training, Cog still behaves like a toddler. It can interact with human beings at the level of a 2 year-old - reacting to sounds and reaching for things. It learns about the world through trial and error, its achievements are reinforced by human researchers, its failures negatively reinforced. Brooks believes that by training Cog the way we train a child, we can develop a thinking machine that will tell us something about ourselves and the nature of thinking and even consciousness.[16] The chasm between CYC and Cog is enormous despite the fact that they are both AI machines.

CYC is a machine full of common sense facts - some of them very humdrum such as "Water is wet" and "Fire is hot". Dennett's argument suggests that it also needs to be programmed with facts like "Fire is not wet." Cog, on the other hand, might find out that fire is not wet when it discovers, through experience, that fire is hot and water is wet. The problem with discovering things through

---

[15] Ibid. p.140-141.

[16] For more information about these and other AIs, see *Time Magazine*, April 1, 1996.

experience is that a robot might get burnt, or short-out its circuits, as it learns to avoid danger. If common sense was acquired through trial and error alone, most of us wouldn't survive our childhood.

AI researchers usually place artificial intelligences in stripped-down versions of reality and set them very specific tasks. For example, a typical AI robot will be asked to move blocks around in a room. The blocks might be labelled and coloured and the robot be required to put block A on top of block B. If it was then asked to move block B into the corner, would it make the classic mistake of supermarket shoppers who remove the can at the bottom of the stack? How does a programmer tell a robot that when a block is on top of another block you can't remove the bottom block? In his paper "Modelling Change: The Frame Problem"[17], Lars-Erik Janlert examines a number of different ways in which this may be done. In each case the robot is told where all the blocks are and what configurations they can be in - this is the robot's model of the world. The AI researcher's chief problem is keeping the robot abreast of changes in its world model - or more precisely teaching the robot how to track the changes itself.

These block worlds are a far cry from the real world which we (and Cog) have to navigate. Such block worlds have very few elements and are designed to be free of surprises for their robot denizens. Some progress has been made in AI by equipping robots with a working knowledge of such stripped-down environments, and getting them to perform limited tasks. The question remains as to whether this block world approach can be scaled up to produce an AI that can act in the real world.

Both Janlert and Lenat believe that common sense knowledge of the world is the key to success, but, as Lenat notes, this common sense knowledge is proving difficult to codify.

> Many of the prerequisite skills and assumptions have become implicit through millennia of cultural and biological evolution and through universal childhood experiences. Before machines can share knowledge as flexibly as people do, these prerequisites need to be recapitulated somehow in explicit, computable forms.
>
> For the past decade, researchers at the CYC project in Austin, Tex., have been hard at work doing exactly that. Originally, the group examined snippets of news articles, novels, advertisements and the like and for each

[17] Lars-Erik Janlert. "Modelling Change: The Frame Problem", in *The Robot's Dilemma: The Frame Problem in Artificial Intelligence* edited by Zenon W. Pylyshyn, New Jersey: Ablex Publishing Corporation, 1987.

> sentence asked “What did the writer assume the reader already knew?” It is that prerequisite knowledge, not the content of the text, that had to be codified. This process has led the group to represent 100,000 discrete concepts and about one million pieces of commonsense knowledge about them.[18]

Note here that Lenat is providing CYC with frame information in exactly the format which Dennett suggests human beings do not store it - in the form of a long list of assertions. Lenat’s method could be characterised as the brute force method of developing artificial intelligence. It requires a vast database and vast computing power. But can we call the process which results *common sense*? Lenat provides an example of CYC’s “common sense” reasoning. CYC is asked to find a picture of a person who is wet. One picture has the caption “Salvador Garcia finishing the marathon in 1994”. CYC reasons:-

Salvador Garcia is a person who has been running for more than two hours.

Salvador Garcia has been doing high exertion for more than two hours.

Salvador Garcia is sweating.

Salvador Garcia is wet.[19]

This is supposed to show that CYC has made a common sense inference from non-explicit evidence. It doesn’t strike me as the kind of process which I would follow given the same task. It also highlights the fact that CYC cannot read pictures - it needs the caption describing the picture. Captions, as we know, are often misleading. It is quite likely that Salvador Garcia will not be sweating in the picture CYC has selected - and CYC won’t know it has failed unless someone tells it. CYC still requires human intervention to keep it on track, telling it where it has erred, and parsing English language requests into its own special version of English - CYC-NL. One problem which CYC’s “knowledge enterers” encountered early on in the project was that CYC found that some of the common sense assertions it was receiving contradicted previous assertions.

> Each assertion in CYC (a statement of fact or a "rule-of-thumb") is located in (or associated with) a specific microtheory or context. Each microtheory captures one "fairly adequate" solution to some knowledge representation area (knowledge domain). These solutions may address general areas like representing and reasoning about space, common devices, time, substances, agents, and causality or specific areas like weather, manufacturing a particular thing, and walking.

[18] Douglas B. Lenat. “Artificial Intelligence” in *Scientific American*. September 1995.
[19] Ibid.

> Different areas may have several different microtheories, since the way an area is perceived or modeled may be different. Different points of view, different assumptions, different levels of granularity, and even what distinctions are important or not important may be significant enough to require creating a separate microtheory. A microtheory may be considered to be a smaller and more modular knowledge-base within CYC, which is specialized on a particular topic.
>
> The important thing to realize is that neither the CYC team, nor CYC itself claims to have a unified theory of time, space, and the universe. Nor does it embody some great master Laws of Thought. What they do have is a suite of specialized microtheories whose union covers the most common cases.[20]

The contradictions between the partitions - or frames, as they are called - are not only at the level of content. In a 1991 memo, Lenat discusses the problem of "divergence" concerning vocabulary.[21] The research teams work in separate groups, each on a particular microtheory, but it was found that the way one group used a term would differ from another.

> Each group enters its micro-theory into a context. Different contexts may use different vocabularies, may make different assumptions, may contradict assertions made in other contexts, etc.....Both knowledge entering and problem solving go on in a context. Axioms external to a context are imported (lifted) from other contexts, using articulation rules. So the question of `what to share' is partially decided at knowledge-entering time, by humans, and partially at inference time, by the system.[22]

What Lenat's colleagues discovered was that even at the level of common sense, knowledge is task specific. Alan Roberts of Monash University believes that this is the crucial issue in AI. He argues that even an "intelligent" robot needs to be programmed by a programmer with data relevant to some purpose.

> [I]t is the programmer, of course, who decides on that purpose, and chooses the data which will be relevant to it. Thus the computer operates in a small, self-contained, relatively unpuzzling world. In contrast, we poor humans have to lumber around in a world capable of infinite novelty and do the best we can with it.[23]

---

[20] "The Unofficial, Unauthorised CYC Frequently Asked Questions Information Sheet." Written by David Whitten

[21] Memo from Doug Lenat via Interlingua Mail, 27th Nov 1991. See Appendix.

[22] Ibid.

[23] Alan Roberts. *Arena Magazine.* Feb-March 1993, pp.34-36. Roberts is a researcher in theoretical ecology. The article is based on a paper entitled "Interventions" delivered at the Monash Craft Conference in 1992.

Roberts' argument leads to a paradox which almost seems to pre-judge the failure of AI. If CYC has to be told what is relevant and irrelevant, what is a meaningful inconsistency and what is a trivial inconsistency, it is not really doing the intelligent thinking. Each night CYC reviews its daily input, categorising it, and forming analogies between various pieces of knowledge. This process sometimes resolves contradictions, and often gives rise to questions which it asks the researchers in the morning. So far this has proven adequate, but Lenat concedes that ,

> the growth of the knowledge base could conceivably outstrip this protection mechanism and allow fatal divergence to set in. "If there are too many inconsistencies," he says, "the knowledge base will collapse." With all its energy devoted to reconciling contradictions, Cyc would lose the ability to do anything else.[24]

Most AI researchers agree that the problem, and the solution, lies in the mode of representation of the knowledge. David Freedman put it thus,

> For example, Cyc needs a better way to distinguish between statements of fact like "Jupiter is the largest planet in the solar system" and statements of opinion like "The Reds are the best team in the National League."[25]

Despite the fact that CYC is designed to overcome this renowned "brittleness" in information systems Marvin Minsky comments,

> "Cyc has a rather logical structure," he notes. "Lenat is trying to make it more flexible with frames, but it's still a single way of representing knowledge, and no one representation will work well. I think the systems of the future will have two or three different ways of representing knowledge with cross-links between them. That's how the brain works: one part has knowledge about people, another about how things work, and so on, to hundreds of specialised areas."[26]

The prevalent view amongst AI researchers is that the solution to the frame problem lies in finding a mode of representing some kind of ontology, or world-view to its creations that is flexible enough to assimilate change.

Brooks explicitly shuns the problem of representation. Because his agent Cog is embodied in the world, in a sense, the world itself is its model. It doesn't need that world to be digitally abstracted for it. In fact, Brooks argues that some

[24] "Commonsense and the Computer" by David Freedman.
[25] Ibid.
[26] Ibid.

aspects of how things are in the world are not digitally abstractable. Brooks approach to developing humanoid intelligence is based on the arguments of a number of theorists, notably G. Lakoff and M. Johnson.

> Their central hypothesis is that all of our thought and language is grounded in physical patterns generated in our sensory and motor systems as we interact with the world. In particular these physical bases of our reason and intelligence can still be discerned in our language as we 'confront' the fact that much of our language can be 'viewed' as physical metaphors 'based' on our bodily interactions with the world.[27]

Brooks and Stein go on to declare that their project proceeds from the above hypothesis. They regard any symbolic abstraction of the world introduced into the robot to be not a statement of how things are, merely a useful way of viewing them - a post-hoc explanation.

> In building a humanoid, we will begin at this sensory level. All intelligence will be grounded in computation on sensory information or on information derived from sensation. However, some of this computation will abstract away from explicit sensation, generalising e.g., over similar situations or sensory inputs. Through sensation and action, the humanoid will experience a conceptualisation of space: "up," "down," "near," "far," etc. We hypothesise that at this point it will be useful for observers to describe the behavior of the robot in symbolic terms ("It put the red blocks together.") This is the first step in representation.
>
> The next step involves a jump from the view of symbols as a convenient but post hoc explanation (i.e. for an observer) to a view in which symbols, somehow, appear to the agent to exist inside the agent's head. This second step is facilitated by language, one of the tools that allows us to become observers of ourselves. This is the trick of consciousness: the idea that "we" exist, that one part of us is observing another.[28]

This manifesto statement demonstrates the depth of the gulf between CYC and Cog. Cog is primarily designed to gather its own data through its body and senses. CYC is not an embodied agent, and although it knows what seeing and hearing are - it can do neither. It cannot learn the way that human beings learn. These approaches to AI development are totally opposed. One is a data-input based model, which emphasises knowledge, the other is an almost evolutionary model which emphasises learning. Both Lenat and Brooks envisage a critical point at which their AI will begin to think. What is surprising is that both identify

[27] Rodney A. Brooks and Lynn Andrea Stein. *Building Brains for Bodies*. MIT Artificial Intelligence Laboratory Memo 1439, August 1993.
[28] Ibid.

the development of the capacity for language within their robots as the key to artificial intelligence. In each case, it is thought that the acquisition of language will mark the point where the AI has become an autonomous agent able to learn for itself.

How crucial is physical embodiment to an artificial intelligence? Is it significant that CYC has no senses? This issue has the potential to split the AI world into two camps - advocates of embodied AIs, and advocates of digitally embodied intelligences. Clearly there are a number of intermediate forms. In science fiction we find computers, robots, androids, cyborgs, replicants, and myriad other forms of intelligences. There are even intelligent space-ships[29]. HAL the computer in *2001 : A Space Odyssey*, can see, hear and talk, and is, according to Lenat, the inspiration for CYC. CYC however, is just a piece of software, downloadable onto any PC. The view that intelligence is something beyond the physical is a pervasive myth. This myth of a disembodied human essence is supported by many religions and many philosophies. It tends to support the view that intelligence need not be connected to a physical form. In science fiction, cyborgs and androids usually have a combined biological and mechanical make-up, and are always more advanced than mechanical robots. On the other hand, energy creatures, and creatures able to exist independently of a fixed body are usually depicted as the most advanced of species. According to Brooks and co. this kind of thinking about intelligence has caused AI researchers to ignore the importance of embodiment. An extreme version of the Brooks argument is that one needs to be *biologically* embodied to develop intelligence.

The inability of researchers to achieve wide agreement about the intelligence of other species which exist on Earth, such as dolphins, and chimpanzees, should be enough to ring warning bells for AI researchers. For decades, biologists and psychologists have been receiving large research grants to teach rats, monkeys and dolphins to do all manner of tricks - navigating mazes and learning language are the favourite tricks being taught. The chief motivation the animals are given is food, and much of the research is Skinnerian - it utilises Skinner's behaviourist stimulus-response methods. Not surprisingly, AI researchers are trying these techniques on their AI agents. Neural net systems have enabled the

development of programs that can learn. Unfortunately, they have not yielded the expected breakthrough in the AI area. The behaviourist idea that human learning is largely a matter of stimulus-response conditioning has been thrown into doubt by these experiments. In theory, a computer could learn English, for example, without being given rules or an algorithm[30]. If you give a learning computer hundreds of examples of sentences, it could then generate its own sentences and be told which were grammatical and which not. This process is repeated again and again until the computer hardly needs to be corrected at all. If successful, this computer will have found a pattern in the sentences which will enable it to generate grammatical sentences. The pattern which it finds may, or may not be the rules of English grammar. This, and similar strategies have reaped some rewards, but a computer which can learn English like a toddler is a long way off. Learning the rules of grammar is one thing, being able to generate sentences that make sense is quite another. Noam Chomsky's famous example - the non-sensical but grammatically correct sentence "Colourless green ideas sleep furiously." - was designed to demonstrate precisely this point. Unfortunately, it was an example which backfired on Chomsky who was subsequently subjected to a barrage of criticism from those who argued that in a given situation the sentence could make perfect sense. We will return to Chomsky and his detractors at a later point. The path from grammatical pattern recognition to sentences which make sense is clearly a thorny one.

Pattern recognition is often considered to be crucial to intelligence[31]. Roberts notes that one of the more surprising findings has been that even non-human animals and non-mamallians use some kind of patterning to shape their perception of "input stimulus." Roberts comments,

> If we look at the experimental findings we will be inclined to ask, not whether a computer could 'learn' like a human being, but rather whether it could ever learn as effectively as a pigeon.

---

[29] See an episode from *Star Trek: The Next Generation* where the Enterprise becomes sentient.

[30] NETtalk, developed by Terry Sejnowski and C.R. Rosenberg, is a heuristic neural net system which is learning to read English text aloud with correct pronunciation. After a certain amount of teaching by English speakers, it generates its own rules and tests them. See *The Quark and the Jaguar* by Murray Gell-Mann for an account of this and other projects concerning complex adaptive mechanisms.

[31] One problem with these learning experiments is that it is difficult to truly say which has identified the pattern, the computer or the researcher.

> We might even ask: will any computer ever have as much *common-sense* as a pigeon?[32]

This argument cuts both ways when we come to consider the type of robot we would like to build in order to make our midnight snack. If we build a robot which learns from experience, we might avoid the frame problem, but go hungry waiting for our snack. If Roberts is right, and computers are incapable of the pattern recognition typical of organisms, we might never eat. Even Brooks is pessimistic that Cog will ever exhibit the robustness and adaptability that seems to characterise even the most primitive biological systems.[33] So we will have to supply our robot, Midnight, with a lot of Lenat type information. Clearly we are considering a kind of combination robot featuring all that is best of bottom-up and top-down approaches.

How much frame information do we need to provide before Midnight can make our midnight snack? What constitutes frame information in this context? If we program Midnight to deal with an empty mayonnaise jar - is that cheating? How much of what we program the robot to do is effectively doing its thinking for it? As long as we continue to use the word “program” as opposed to “train” it will always seem as if the thinking is being done by the programmer. If we assume that Midnight can respond to hints spoken in English which we volunteer when it encounters a problem, this seems more like training than programming. Crucially we would have to have access to some of Midnight’s internal workings to judge whether it actually made a decision and did some thinking. This problem is not one peculiar to AI. It is a recognised educational problem that if a student submits an essay to a teacher time after time until it is deemed OK, it is not always possible to say who has done the thinking - the student or the teacher. If it is an issue with human training, it is also going to be an issue in AI. If the young student were being “taught” to make a midnight snack, I suspect that there would be a lot of data being supplied of the order, “mind that knife - it’s sharp”, “don’t put your hand on the hob - it’s hot”, “you’re going to have to soften-up that butter”, - humdrum facts which nevertheless can influence the outcome of the exercise

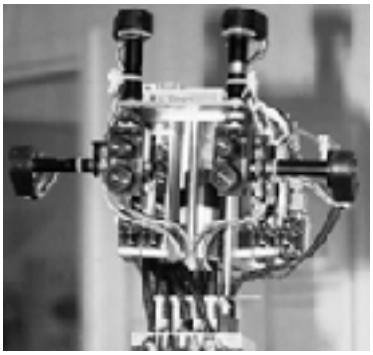
---

[32] Roberts. pp.34-36.

[33] See “From Earwigs to Humans” by Rodney A. Brooks, MIT Artificial Intelligence Laboratory, brooks@ai.mit.edu, 1996 for a review of progress on Cog.

This way of providing frame information to children seems rather haphazard, but it seems to work. Another way is through formal education which tends to be more structured. The combination of experience in the world, interaction with other beings, positive and negative reinforcement and being fed lots of facts, works for human beings. The problem for AI researchers is how to duplicate this process in robots. As Dennett suggests, the backstage story of how human common sense reasoning is achieved is yet to be told[34].

We are already starting to envisage a robot far more advanced than any that has been built, or is even on the drawing board. Luckily science fiction writers have been using robots for decades. Science fiction robots do everything from flying spaceships (e.g. HAL in *2001*) to domestic chores (e.g. Kryten in *Red Dwarf*). They can be fighting machines or pleasure machines (e.g. the replicants in *Blade Runner*), god-like (Deep Thought in *Hitchhiker's Guide to the Galaxy*) or completely stupid (Holly in *Red Dwarf*). Some of them even have emotions (Marvin the paranoid android in *Hitchhiker's Guide to the Galaxy*, and Data in *Generations*). Not surprisingly, this vast spectrum of robots and computers has anticipated many of the AI problems we have discussed and some that AI researchers haven't even begun to deal with. I should note here that Number 5 bears an uncanny resemblance to Cog.



The configuration of "eyes" and arms is very similar. I have no idea what the direction of influence was here. I do know that the influence of science fiction on artificial intelligence research is profound. It is a source of inspiration in research and teaching, and models problems in ways which make them accessible to non-specialists. The problems which science fiction deals with are usually more philosophical than technical, and often introduce a moral dimension. *Short Circuit,* for example, not only dramatises the frame problem, but looks at the

[34] "Cognitive Wheels" p.148.

consequences of solving the frame problem and facing the ensuing ethical problem of a free-thinking robot.

The problems of building an intelligent, embodied, robot with the advantages of both logical and common sense reasoning, specialist knowledge and everyday knowledge, flexibility and robustness, have been explored with wit and imagination in a host of science fiction stories, and these explorations will serve as stepping stones in our midnight snack thought experiment.

Logic and Common sense: Is it possible to program a robot with common sense?

Lenat describes the expert systems which AI has so far developed as “idiot savants”

> Ask a medical program about a rusty old car, and it might blithely diagnose measles.[35]

Furthermore, it wouldn’t know it had made a mistake. It made a perfectly logical deduction, but lacked the common sense information that cars don’t get measles.

It is this “brittleness” of information systems which Lenat originally set out to overcome. Numerous science fiction stories contrast logic and common sense. In many instances the relentless logic of the computer mind is both its strength and its weakness. The Terminator, for example, performs its programmed task without ever hesitating - its relentlessness in a human being would be considered obsessional. Human beings can only outwit it because its programming makes it predictable.

Science fiction stories often suggest that beings which operate on logic alone find themselves compromised by their adherence to it. The classic story depicting the limitations of logic is Gordon R. Dickson's "The Monkey Wrench." On a weather-station in sub-zero temperatures on Venus, two men make a wager as to whether the computer "Brain" which is responsible for maintaining the weather-station can be made to malfunction. Cary Harmon, the lawyer, bets Burke McIntyre, the meteorologist, a year's salary that given one minute at the speech interface he can render the machine out of order. At the end of the story,

[35] Douglas B. Lenat. “Artificial Intelligence” in *Scientific American*. September 1995.

we are left with the image of the two men awaiting death in the freezing and darkened weather station as the Brain dedicates its whole effort to solving the paradox fed to it by Cary. The paradox which he feeds the machine is a classic riddle,

> "You must reject the statement which I am now making, because all the statements I make are incorrect."[36]

Much of logic involves ascribing truth values to statements. When the computer attempts to ascribe a truth value to the above statement it runs into a logical loop. If the statement is accepted as true, it must be false. On the other hand if it is accepted as false, then it must be true, and so on. The computer's attempt to solve this problem involves it in a process which locks it into a loop which prevents it from devoting any time to maintaining life support. Lenat is worried that CYC might undergo a similar breakdown if the weight of contradictions in its knowledge base becomes so great that it needs to spend all its time resolving them. “The Monkey Wrench” is primarily a warning about dependence on mechanised systems, but it also counterpoises logic and common sense. The problem of circularity in logic and mathematics was found to be a symptom of a more serious problem when, in 1931, Kurt Gödel established that it is impossible to prove that all the propositions of arithmetic can be inferred from a finite set of consistent axioms.[37] If one makes a series of statements which one knows to be true within a system, at least one of those statements cannot be proven within the system. The truth of at least one of the propositions must be proven from outside the system. Alan Turing, one of the pioneers of AI, made a similar discovery. Like mathematics, the operations of computers are based on algorithms - mechanical operations. Turing established that some algorithms were non-computable - that a machine which ran such an algorithm would continue calculating without ever arriving at a solution. The ability of mathematicians to choose algorithms which have solutions, is outside the capacity of machines which merely run algorithms. The implications of Gödel’s proof and Turing’s experiment are far-reaching for mathematics, logic and

[36] Gordon R. Dickson. "The Monkey Wrench" reprinted in *The Penguin Science Fiction Omnibus*, Harmondsworth: Penguin, 1973, p.212.
[37] Kurt Gödel. *On Formally Undecidable Propositions*, translated by J. van Heijenoort, in *From Frege to Gödel: A Source Book on Mathematical Logic*, 1879-1931, ed. by J. Van Heijenoort. Cambridge, Mass: Harvard University Press, 1967.

seemingly for AI, because it establishes that mathematics is not self-evidently complete, and neither is a system, such as a digital AI, which depends on it.

Human beings have the capacity to break a logical loop using “insight” or “intuition” and this seems to set them apart from merely logical machines. In a sense, human beings partially pre-solve problems by choosing ways of solving problems which are more likely to lead to solutions. This ability to choose the appropriate course of action from the plethora of possible courses of action is precisely the ability we would like our robot to have. Opponents of AI have suggested that such capacities are a characteristic of the conscious mind and completely beyond the capacity of a machine that “just runs programs”. Dennett disagrees and argues that Gödel’s proof merely discounts the possibility that any given system in mathematics can prove all its own axioms. Something which a computer, for example, rarely needs to do because a computer system can arrive at perfectly good solutions to mathematical problems, chess problems, and a host of other problems, without needing to develop a self-sustaining proof.[38]

The issue is dealt with very deftly in an episode of *Star Trek: the Next Generation* in which Data is beaten at a game called Strategema by a grand master.[39] Data consequently withdraws from his duties as first officer. His reasoning is that if he can be beaten in a game where he didn’t make a mistake, he can be wrong in decisions he makes on the bridge, and thereby unwittingly endanger the ship. Captain Picard tries to convince him that losing at Strategema does not affect his value as a crew member. He argues that Data is still able to perform all the functions he usually performs despite the fact that he lost at Strategema. Unfortunately Data reasons that he can’t - in a sense Data is in a logical loop. Anyone who uses a computer knows that machines often “lock-up” when they find themselves in logical loops. It may be that the frame problem and the problem of logical circularity are linked. Both would be solved if the computer/robot could break out of exhaustive logical analysis of the problem or the situation. In effect, it would be useful if it simply got bored.

---

[38] Daniel C. Dennett. *Darwin’s Dangerous Idea*. Harmondsworth: Penguin, 1996.

[39] *Star Trek: The Next Generation*, “Peak Performance", written by David Kemper, 1990. Data’s decision to suspend himself from the bridge could be viewed as a faulty component in a machine being declared faulty (declaring itself in this case) and being replaced.

The conflict between logic and humanity is a familiar theme in science fiction. In a scene from a *Next Generation* episode entitled "Unification", an interchange between Spock (half Vulcan, half human) and Data (the android) cleverly highlights the issues. Data tells Spock that he has selected Picard as his role model in his "quest to become more human." Spock is "fascinated" that Data should strive to become "more human" when Data already possesses the characteristics which most Vulcans strive for all their lives. This scene establishes the crucial differences between two characters who hitherto one might have been tempted to twin. In their respective series, both function as paragons of logical thinking in situations which are often dominated by human emotions. From this perspective one is tempted to suggest that, dramatically, the characters serve similar structural functions. In fact, as the scene shows, they are opposites. Spock strives to suppress his humanity in favour of a logical Vulcan way of life. Data strives to be more human. Data remarks to Spock, "In effect, you have abandoned what I have sought all my life." and then Data asks Spock if he has "missed his humanity?", Spock responds that he has "no regrets". A comment which Data characterises as a "human expression". This exchange seems to put Data out in the cold. Spock is a being with choices which Data, as a machine, does not have.

Data's quest involves him trying to understand human behaviour such as lying, falling in love, and humour. There is a whole episode of *The Next Generation*[40] largely devoted to Data's quest to understand comedy. This episode is reminiscent of the time when Data decides that he needs to learn to dance. He faultlessly copies another dancer's movements and quickly becomes a superb tap dancer. The visual cues in this episode constantly undermine Data's performance as a dancer by suggesting that his performance is mere imitation; he is made to look like a marionette.[41]

[40] *Star Trek: The Next Generation*, "The Outrageous Okona", teleplay by Harold Aptor and Ronald D. Moore, 1990.
[41] *Star Trek: The Next Generation*, "Data's Day", teleplay by Harold Aptor and Ronald D. Moore, 1990.

Data fails in his quest to become a stand-up comedian because, although he can easily mimic great stand-up comedians, he does not know when things are funny, consequently he cannot come up with a new joke, or even recognise a new joke. Humour, it seems, is not something one masters through logic. Neither it seems, is poker.

Digital Experience

"The Measure of a Man" begins with Data declaring that the game of poker is "exceedingly simple, with only 52 cards, 21 of which I will see, and four other players, there are a limited number of combinations."[42] It is tempting to think that a calculator with total recall would make a very good poker player. In the game that ensues both Data and Riker appear to have winning hands, but when Riker ups the stakes, Data folds. Data is amazed to discover that Riker had nothing - Riker was bluffing. Later he describes the experience to Colonel Maddox who wishes to disassemble Data so that he can study Data's design and build more Datas. Data is reluctant to have his memory downloaded into a mainframe, and argues that he is more than the sum of his circuits and memories.

Data: Reduced to the mere facts of the events - the substance, the flavour of the moment could be lost. Take games of chance
Maddox: Games of chance?
Data: I have read and absorbed every treatise and text book on the subject and found myself well-prepared for the experience. Yet when I finally played poker, I discovered that the reality bore little resemblance to the rules.
Maddox: And the point being
Data: That while I believe it is possible to download information contained in a positronic brain, I do not believe you have acquired the expertise necessary to preserve the essence of those experiences.

[42] "The Measure of a Man" *Star Trek: The Next Generation* written by Melinda M.Snodgrass and directed by Robert Sheerer, Paramount Pictures 1989.

There is an ineffable quality to memory which I do not believe can survive your procedure.

Maddox: Ineffable quality.[43]

The word "ineffable" is remarked by Dennett as a favourite of philosophers wishing to distinguish the quality of experience (so called "qualia") from the facts of an experience.[44] Data is saying that Maddox's downloading procedure is not adequate to the task of extracting and preserving the dispositions and qualities of experience which Data has accumulated. In other words, Data's files on the game of poker may be downloaded, but his surprise at finding he had been bluffed cannot be digitalised. Can surprise be digitalised? Put so baldly, one see why some philosophers hang onto the idea of private, phenomenal properties of experience.

The mystery here is not as deep as it seems. CYC is not embodied, therefore it cannot learn about cars by going for a drive in the country. Cog on the other hand can be taken for a drive and respond to sights and sounds and movement just like a child. Clearly, the two machines are going to arrive at entirely different views of what cars are. One will be a digital representation of lots of knowledge about cars, the other a digital representation derived from sensory experience of cars. Brooks would agree with Data. He claims that it is not possible to adequately digitalise much of the analogue input which Cog receives. He argues that digital abstraction "hides the details of perceptual and motor processes."[45]

> By hiding details of analog voltages that constitute our systems, the digital abstraction facilitates reasoning about and construction with these elements.....certain portions of the resulting system may never be interpretable in terms of the digital abstraction.[46]

Brooks doesn't believe that Data could be digitalised because parts of perceptual and motor systems are inherently undigitalisable. The question which remains is whether these parts are crucial to what we call intelligence.

Data's predicament serves to warn us that even with all the information Midnight needs about midnight snacks - recipe books, cooking programs etc., it might still fail - just as Data failed in the game of poker. Data's vast knowledge combines

---

[43] "The Measure of a Man"

[44] Daniel Dennett. *Consciousness Explained*. Harmondsworth: Penguin, 1991, pp.49-50.

[45] Rodney A. Brooks and Lynn Andrea Stein. *Building Brains for Bodies*. MIT Artificial Intelligence Laboratory Memo 1439, August 1993.

downloaded information and information gleaned from experience, he is therefore a sort of combination AI robot. He was raised as a virtual child for several years (like Brooks' Cog), but his learning was supplemented by feeding him gigabytes of data about the world (like Lenat's CYC). This combination enables him to navigate most everyday situations, and a lot of very bizarre ones, with relative ease, yet Data remains puzzles by much human behaviour.

Data is not the only science fiction robot with failings when it comes to understanding human behaviour. Science fictional robots often have gaps in their data banks when it comes to common human behaviour, such as crying (see *Terminator 2* [47]), or eating (*Short Circuit*). In such stories the limitations of logic are highlighted in order to emphasise the infinite subtlety and variety of human creativity.[48] Number 5 has assimilated gigabytes of "facts", but it is clear in the kitchen scene, that it hasn't developed the creative skills to apply them. Later in the film, Number 5 is shown to have transcended its programming and is recognised as an intelligent and creative being.

Creativity

In science fiction, creativity[49] is often used as a marker of intelligence. In *Star Trek: The Next Generation* we often see Data painting; Number 5 sees butterflies in ink blots; in Asimov's "Bicentennial Man," [50] the robot called Andrew Martin, begins his path to humanity through woodcarving. Science fiction stories involving robots and computers are often a meditation on the nature of creativity.

---

[46] Ibid.

[47] In *Terminator 2: Judgment Day* the terminator played by Swarzenegger asks John Connor "why do your eyes water?" It is curious that the robot knows so much about human life but has no data on crying.

[48] Robert Silverberg explores this theme in his 1956 story "The Macauley Circuit". Reprinted in *Machines That Think* edited by Isaac Asimov, Patricia S.Warwick, and Martin H.Greenberg, Penguin, 1983

[49] The nature of creativity is the subject of wide debate. Edward de Bono, for example, suggests that it is linked to what he calls "lateral thinking". Others say that it is determined by which side of the brain one uses, or that it is marked by the ability to perceive patterns. Perhaps the ability to navigate the everyday world, the ability which we call common sense, is inherently creative. One can assimilate as many facts as one likes about the world and the things in it, but knowing when and how to act on these facts seems to require a whole new level of thinking. On a hot day, pigeons stay out of the sun. However many human beings haven't got the "common sense" to do this, yet I don't know of anyone who regards pigeons as creative. In this instance the pigeons are acting on instinct - they are "born knowing" (as Dennett might say), that sitting in the sun is dangerous.

[50] Isaac Asimov. "The Bicentennial Man." rpt. in *Machines That Think*. ed., Isaac Asimov, Patricia Warrick, and Martin Greenberg. Harmondsworth: Penguin, 1986, pp.495-519.

Robots are often granted a measure of common sense, but creativity is rarely granted (*Short Circuit's* Number 5, and Asimov's "Bicentennial Man" are notable exceptions). Even the sophisticated Data is not granted unqualified creativity. In the episode "Elementary Dear Data" it is suggested that Data's deductive reasoning is inferior to the kind of reasoning which characterises the deductive reasoning of Sherlock Holmes. One of the *Enterprise*'s doctors, Dr. Pulaski, argues that Data could never solve a genuine Holmes mystery which he hadn't read, because he lacks understanding of the "human soul."

> Dr Pulaski (*to Geordi*): Your artificial friend doesn't have a prayer of solving a Holmes Mystery he hasn't read.

In response to this challenge, Data quickly solves a composite Holmes mystery written by the computer and played out in the computer-generated, hologrammatic 'reality' of the "holodeck"[51]. Despite Data's success, Dr Pulaski remains unconvinced that Data has solved the mystery using the kind of insight with which Conan Doyle imbued Holmes. Data defends himself in the following terms,

> Data: Reasoning from the general to the specific. Is that not the very definition of deduction. Is that not the way Sherlock Holmes worked?
>
> Dr Pulaski: Variations on a theme. (*to Geordi*) Now, now do see my point? All that he knows is stored in his memory banks. Inspiration, original thought, all the true strengths of Holmes. It's not possible for our friend. (*to Data*) I'll give you credit for your vast knowledge, but your circuits would short out if you were confronted with a truly original mystery. [52]

What Dr Pulaski is asserting here is that no matter how vast Data's memory banks are, Data's actions and thinking amount to no more than sophisticated imitation. Similar arguments are often used by those hostile to AI. John Searle, for example, argues that computer simulations of thinking are no more like thinking than computer simulations of storms are wet.[53] The more extreme

---

[51] The holodeck is an innovation in *The Next Generation* which allows the crew of the enterprise to set up interactive computer simulations of almost any environment. The simulation is a hologrammatic environment which enables interaction with a whole range of characters programmed into the computer and brought to virtual life. In an article in *Omni* (Dec. 1994) the designers of *Star Trek* admit that the conceptual parameters of the holodeck are rather loose. In the first section of *Consciousness Explained* Dennett argues that such environments could never be built.

[52] "Elementary My Dear Data" episode of *Startrek: the Next Generation.*

[53] John Searle, Reith Lectures, 1978.

version of this view maintains that mere physical systems of electrical and chemical interactions cannot produce something as mysterious as consciousness. In many of these arguments it is not always clear what is missing from the imaginary robot. In the above episode, a holodeck adventure follows in which the Enterprise's crew is pitted against a hologrammatic Moriarty who gains control of the holodeck computer. The doctor and the ship are endangered, but it is not the super-android Data who saves the day, but the very human Captain Picard. In effect, Dr Pulaski wins her challenge concerning Data's creativity. In the event, it is an unsatisfactory conclusion because the story doesn't clarify what Data lacks. Lacking frame information, lacking common sense, lacking consciousness, and lacking experience might all turn out to be entirely separate problems. On the other hand, they might be inextricably linked. Data's failure at poker may have resonances of similar failures for many human beings. How many times have we prepared for a new experience, like sailing or bungy jumping, only to find ourselves overwhelmed - or at least surprised - by the reality?

Surprise: Frame Problems and Human Beings

The television series *Mr. Bean* involves an unfortunate man who cannot seem to cope with all the mundane things we do every day. Shaving, driving, cooking, going to a restaurant, are all trials for him because he lacks frame information. Throughout the series we see him invent remarkably novel ways of negotiating situations which we deal with almost without thinking. When he makes a mess of eating a sandwich, or throwing a party, it highlights how complex these activities are. In one episode he throws a New Year's Eve party, but his friends get so bored that when he is out of the room they move the clock forward an hour so they can sing the New Year in and slip off to another party. Mr. Bean is so gullible he falls for the ruse and is in bed by midnight. He takes everything at face value. Normal human beings, on the other hand, are constantly questioning things. I was having lunch at a restaurant recently and looking at the restaurant clock was surprised to find it was only 11.30 a.m.. I remarked to my companion that the restaurant clock was wrong. Then I thought about it being Sunday and November and how easy it had been to get a table, and eventually asked the

people on the next table whether the clocks had gone back for daylight saving. This kind of suspicious questioning, and putting together of evidence from disparate sources, is typical of human thinking. Computers can't suspect things. Gary Kasparov comments on this difference between human and computer intelligence in a short account of his first encounter with IBM's Deep Blue chess computer. The computer (white) defeated Kasparov in game 1, using an early pawn sacrifice which fractured black's pawn structure. Kasparov sensed a new kind of intelligence at work - but did some research before game 2, and discovered that the machine was able to compare each position with an enormous database (looking at several million positions per second) 12 moves in advance - something no computer had ever been able to do before. Kasparov notes that the computer didn't view the pawn sacrifice as a sacrifice at all. It recovered the pawn six moves later. This clue was enough to give Kasparov the advantage in the remaining games.

> I was able to exploit the traditional shortcomings of computers throughout the rest of the match. At one point, for example, I changed slightly the order of an opening sequence. Because it was unable to compare this new position meaningfully with similar ones in its database, it had to start calculating away and was unable to find a good plan. A human would have simply wondered, "What's Gary up to?" judged the change to be meaningless and moved on.[54]

Kasparov describes Deep Blue's intelligence as "weird" and inflexible. Crucially, it lacked the ability to suspect that Kasparov was making false moves. He comments that if it "understood" the game, instead of merely searching for material advantage in the game, it wouldn't have been fooled. This gullibility is typical of computers, Mr. Bean, and of course, aliens. In the opening sequence of the *Mr. Bean* series Mr. Bean falls out of the sky in a pool of light - as if he had been dropped from a space-ship - like an alien. The television series *Third Rock from the Sun* exploits a similar idea with the premise of a group of aliens in human form investigating Earth culture and reporting their findings to their superiors. There is a scene in one episode where all the aliens are sitting in a car listening to the lottery results on the radio. As the numbers unfold they get more and more excited as it is revealed they hold the winning ticket - their commander promptly tears up the ticket and says "Its amazing how much fun you can have

[54] *Time Magazine*, April 1, 1996.

for just a dollar." How can these aliens understand human society if they haven't got a concept of money or lotteries? Dennett observes that money is a complex concept which, though real to almost everybody on the planet, is nevertheless only a concept. If we all forgot what money was overnight - there would be a lot of meaningless bits of paper and metal in our pockets.[55] Furthermore, the money concept only works because of human money behaviour. Science fiction depicts robots and aliens as creatures lacking frame information and consequently prompts readers and viewers to think about their everyday assumptions about themselves and their world. In the television show *Sliders* the protagonists "slide" across dimensional boundaries between alternate Earth's. Each Earth is crucially different from the one we know - the Axis powers may have won the 2nd World War, justice might be a lottery, murder might be legal, Kennedy might have lived. Stripping away frame information is one way of making people aware of their assumptions, another way is depicting creatures whose frame information is different from that of "normal" human beings. However, it isn't just aliens and robots whose "frame" is different from other peoples. The popularity of travel documentaries and wildlife programs is testimony to the fascination of exposing people to their hidden (frame) assumptions. Different life styles and different religions, often seem strange to on-lookers. From the outside they often seem to involve a lot of rather hard to swallow assumptions about how the world is.

The series *X-Files* capitalises on the fact that its hero, Mulder, factors in information which most people dismiss as nonsense. Voodoo, alien abductions, and re-incarnation are admissible as explanations for the crimes which Mulder and Scully attempt to solve. The show highlights the fact that we often reject evidence because it doesn't fit how we are supposed to see the world. Many of the paranormal events in *X-Files* defy science - nevertheless Mulder and Scully apply scientific procedures. You may believe that space-ships have visited the Earth. You might even believe that we were actually seeded by space-faring creatures and that they return every now and then to see how their experiment is going - abducting a few people along the way. Does this mean you are crazy? Were the people who panicked when they heard Orson Welles' *War of the*

[55] *Consciousness Explained*. p.24.

Worlds broadcast crazy? There is a character in the film *Independence Day*[56] who is a drooling drunk, having lost the respect of his family and the community because he claims to have been abducted and experimented upon by aliens. By the end of the film he is proven to be right, and this apparently crazy view of the world has been confirmed.

Are aliens part of your frame information? The existence or non-existence of aliens may be a matter of no importance to you. Perhaps you have no view on the matter. Perhaps you accept that the universe must contain myriad life-forms but, because you don't expect to meet any of them, you don't think it is a very important issue. On the other hand you might have encyclopedic knowledge of, and very strong views on football. In your model of reality the alien part is less well modelled than the football part. It is really a matter of priorities. A certain amount of specialisation enables us to negotiate the world quite adequately - except of course when something out-of-the-ordinary happens. The real AI robots we have talked about are all very specialised - the science-fictional ones less so. AI researchers provide models of the environment and situations which the AI is going to encounter. The "original frame problem" according to Janlert is the problem of representing change. This also entails representing what does not change. He notes:

> Whereas it is logically possible that turning on the light in my living room will make the walls turn black, halt inflation, and raise the melting point of lead, nothing of this does in fact happen. [57]

AI researchers find themselves calculating a lot of non-effects and being drowned in a flood of pseudo-laws. A little common sense would cut down all this calculating work and point the AI at the relevant parts of the situation. The general frame problem, Janlert notes is the "problem of finding a representational form permitting a changing, complex world to be efficiently and adequately represented."[58] He suggests that the "operative metaphysics" which human beings use is common sense, and that common sense AI models would have the advantage of being based in the conceptual world of human beings and consequently better placed to communicate with them.

---

[56] *Independence Day*, 1996. Ref?
[57] "Modelling Change" p.6.
[58] Ibid. p.32

Everyone seems to agree that a bit of common sense would dissolve the frame problem overnight. But if your common sense view of reality includes a belief that we are part of an intergalactic genetic experiment, or that Elvis lives, or that Jesus saves, we may have a dispute about which common sense model is the best to use. Janlert notes,

> What goes under the name of "common sense " in AI is mostly a curious mixture of Aristotelian and Newtonian mechanics, not strikingly commonsensical, and not to be uncritically accepted as the operative metaphysics.[59]

Our education provided us with a model of reality which tells us how things exist and move in the world. Whatever it was called, it tended to confirm most of what we already knew about objects falling, and what happened when we heated them up etc.. Many of us forgot most of this science, and our common sense view of the world contains pretty much what we knew before, with a garnish of Newtonian theory. The crucial feature of our common sense model is that it enables us to foreground the important stuff and ignore peripheral problems. When we engage the world we organise the model according to what we see as relevant in that particular situation.

Our three AI researchers, Brooks, Janlert and Lenat, each raise the issue of representation in AI.

Brooks maintains that symbolic abstraction can be a,

> crucial tool in the analysis and synthesis of our humanoids; but we do not necessarily expect these symbols to appear explicitly in the humanoid's head.[60]

Janlert maintains that the frame problem can be resolved by finding a "form of representation" which combines the stability of a "world-version" with the freedom of shifting perspectives.

Lenat rejects the possibility of finding a unified model of the world for CYC, and is inputting knowledge about hundreds of different frames, or contexts (which he calls microtheories), each of which has a basis in common sense.

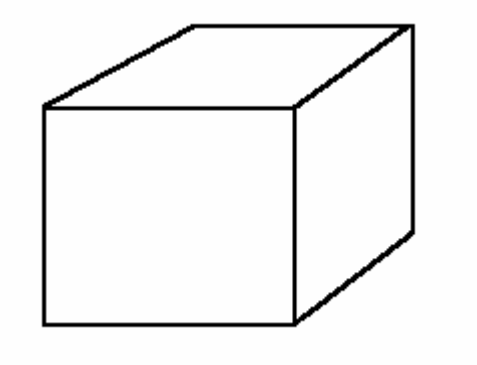
These three views on the issue of models and representation encapsulate three approaches to resolving the frame problem. Another way of clarifying the issue

[59] Ibid.
[60] Brooks and Stein. "Building Brains for Bodies" .

may be to examine how human beings plan their engagements with the world. Perhaps an examination of how human beings develop and use models will help resolve the frame problem.

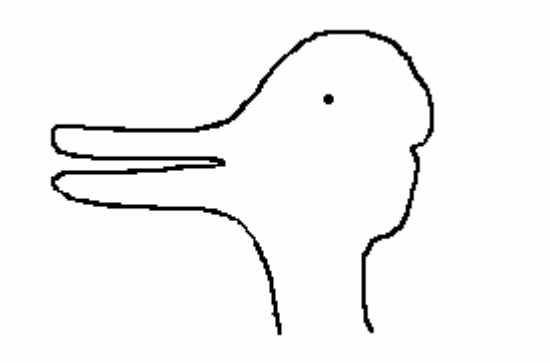
2. Frames and Models: Do we use a model of reality to plan our engagements with the world?

Do we use a model of reality whenever we engage the world? Are smelling, tasting, seeing, hearing and touching modulated by our individual world-views? In *Philosophical Investigations,* Wittgenstein argues that the apparently simple act of seeing involves projecting a kind of model of the world. His argument teases out the differences between seeing, and seeing an aspect of something.[61] He argues that seeing a drawing of a cube as a cube, is very different from seeing a lot of marks on a piece of paper.



Furthermore one can regard the above drawing as three boards edge to edge, as a wire frame of an angle, or as a glass cube. Is it possible to have the above image in your mind, but not be sure which of these three things it is? Does the idea of having an image in mind make sense? An act of interpretation or recognition takes place as we look at the drawing. Wittgenstein also uses Jastrow's famous duck-rabbit example to demonstrate how aspects can change.

[61] Ludwig Wittgenstein. *Philosophical Investigations*. Ref?



One moment we see it as a rabbit, and then as a duck. Wittgenstein goes on to discuss what happens when one sees an object which one doesn't immediately recognise - in bad lighting for example. There is a period where one tries to make the object coherent - there is a very strong drive to see the object as something - and not just a nameless blob.

Daniel Dennett suggests that this "epistemic hunger" to gather information from the world is a crucial feature of consciousness. Primitive organisms, he argues, have simple withdrawal and approach responses to stimuli depending on whether the thing sensed is dangerous or edible.

> In the beginning, all "signals" caused by things in the environment meant either "scram!" or "go for it!" [62]

These primitive creatures are primarily what Richard Dawkins would call "replicators" - their sole purpose is to reproduce themselves. Dennett goes on to explain,

> Put more anthropomorphically, if these simple replicators want to continue to replicate, they should hope and strive for various things; they should avoid the "bad things" and seek the "good" things. When an entity arrives on the scene capable of behaviour that staves off, however primitively, its own dissolution and decomposition, it brings with it into the world its "good". That is to say, it creates a point of view from which the world's events can be roughly partitioned into the favourable, the unfavorable, and the neutral. [63]
>
> Faced with the task of extracting useful future out of our personal pasts, we organisms try to get something for free (or at least at bargain price): to find laws of the world - and if there aren't any, to find approximate laws of the world - anything that will give us an edge.[64]

These organisms could be said to be using a model of reality to navigate the world. They are not just seeing the world, they are seeing it as dangerous,

[62] Daniel Dennett. *Consciousness Explained*. Harmondsworth: Penguin, 1991, pp.177-78.
[63] *Consciousness Explained*. pp.173-174.

delicious, uncomfortable etc.. Dennett suggests that finding laws of the world gives organisms, from primitive cells to complex human beings, an edge in the fight to survive. An organism with such a model of the world hard-wired into it, however primitive, navigates the world through instinct. Its primitive model enables it to survive. A fish sees the world in terms of food, predators and mating. Social animals have to develop more complex world-views because surviving in the social unit is a pre-requisite to surviving in the world. Primates, for example, must be able to recognise the relation and position of others in their group, and behave accordingly. There is some evidence that the larger brains of chimpanzees, for example, evolved partly in order to cope with the complexity of these social rituals, pecking orders and family allegiances.[65] Language enables human beings to represent to themselves possible scenarios, models of the world that are more than just on-the-spot reactions to circumstances. Language enables pre-meditated actions and reactions. Already we can talk about different kinds of models - primitive action/reaction models, social imperative models, and models used as a basis for planning. Perhaps there are more kinds of models operating in the human species, but at this stage we need to be clear what the role of each of the above is, how it operates, and how, in certain circumstances, one model may over-rule another.

John McCrone traces the development of the human mind in his book *The Ape That Spoke*. He suggests that raw pain, pleasure and arousal may be considered innate, and everything else learnt. The pleasure we feel when we eat, the pain when we stub our toe, the arousal we feel when our heart beats faster - these are the carrot, stick, and motor which we are born with. Everything else is learnt by associating it with these three emotions - including language. He describes a process where everything that is seen, smelled, heard, touched or tasted is associated with these emotions and with other things in any given situation. Thus, for a cat, the sight of a mouse activates memories of previous mouse encounters and spurs it into mouse stalking activity. These sights, sounds and smells are linked together in the brain in what McCrone calls

---

[64] Ibid. p.178.

[65] John McCrone. *The Ape That Spoke: Language and the Evolution of the Human Mind.* London: Picador, 1990.

associative nets. Thus when a cat sees a movement something like the movement of a mouse, the net in its brain activates and the cat strikes a stalking stance. Learning, in McCrone's view, is building up these nets of association in the brain so that a sight or sound, or even a word, can, rouse all kinds of associations and spur a creature to action.

This ability to recognise familiar things in their environment - predators, kin, food etc. gives animals obvious evolutionary advantages. McCrone notes,

> long-term memory has the second equally important job of giving animals an internal 'mental backdrop', so that all new experiences can fit into a context of understanding. For instance, a mouse builds up a general picture of the world as it grows up so that every day that passes, its pool of knowledge increases about how corn smells, predators look, or branches bend. This inner picture of the world means that when a problem crops up - such as getting an ear of corn - the mouse already has background knowledge about swaying stems and swooping owls.[66]

McCrone maintains that it is the building of bridges between nets of associations when problems present themselves in the world, that constitutes natural thought. He cites the example of an experiment with young chimps left in a cage with a stick and given the problem of getting hold of a banana which was out of their reach.

> The chimps had never had a chance to play with sticks before and did not attempt to use the one in their cage to hook the food nearer. They tried stretching through the bars to reach the banana a few times and then gave up. However, another set of chimps who had been given three days to play around with a stick before the test saw the answer within twenty seconds. Their apparently aimless play had allowed them to build up a rich net of information about sticks[67].

If our artificial intelligence robot could utilise this kind of thinking, all the knowledge we pump into it would turn out to be very useful when it encountered an unforeseen problem. Unfortunately, we find that these robots can't even recognise a tricky situation - like Deep Blue playing Kasparov, they blunder on regardless of the new situation. This ability to recognise "tricky situations" is not standard equipment for all animals.

> Very simple animals, like worms and jellyfish, can hardly be said to have any thoughts at all. Thought only has meaning for an animal when it can hold some internal representation of the world in its head, making the sort

[66] Ibid. p.81.
[67] Ibid. p. 100.

> of mental maps that lead to consciously felt nets. So worms and jellyfish do not really think. Their nervous systems simply react and adjust to the world in hard-wired fashion. Moving on to fish and reptiles, we see signs of intelligent minds at work, although their cold-blooded metabolism puts a limit on how much they can invest in energetic brain work - or how much they could do about tricky situations if they had the wit to realise they existed.[68]

The process which McCrone describes, whereby associative memories are used to combat novel situations, is not well understood. For the purposes of building a robot, it is precisely this problem of getting the robot to rouse the right memories in the right situations that is the stumbling block (sometimes literally). Brooks admits that Cog's abilities to find its way around the building are insectlike, it doesn't rouse any memories to deal with situations. It could not be said that Cog had an internal representation of the building. It finds its way around by recognising "homeward markers". Cog doesn't face the problem of having to use a model. McCrone suggests that even lizards and mice use an associative process to build a kind of representation of the world in their brains. This representation enables on-the-spot reactions to situations. Without language, this way of thinking is always in the present, and triggered by pain, hunger, thirst or the environment. It is language which enables human beings to break free of the "tyranny of the present".

Planning Ahead

> Humans may have developed language and discovered new ways to use the brain's memory surfaces, but we still rely heavily on the broad understanding of life that natural memory gives us. This is the sort of knowledge about how branches bend and rocks fall that forms the internal backdrop of our thinking. As we have seen, this inner backdrop of common-sense knowledge is so basic that we take it for granted and forget it has to be learned before going on to more advanced problem solving. For example, babies need to learn that things still exist even when they disappear from sight - like when a mother hides a rattle behind her back -[69] .....

This knowledge which forms the 'mental backdrop' for thinking is largely pre-linguistic according to McCrone, and corresponds with what we have called

[68] Ibid. pp. 99-100.

"frame information". According to some sources it is also pre-logical. There is some evidence to suggest that deductive reasoning cannot develop in children until this common sense reasoning is established.

> It is a great myth of modern man that we are inherently rational. We talk about deductive logic as if it were wired into our brains and were the main difference between us and other animals. But formal logic is a very recent creation of modern man. The artificiality of Western logic is highlighted by the navigation feats of the Truk islanders in the South Pacific who regularly sail hundreds of miles between little coral islands, finding their way by 'feel'. The trained Western navigator would find his way around by charts and measurements. If asked at any time where he was, he could point to the map and give a logical step-by-step account of the course he must follow to get to his destination. The Truk islander on the other hand has a mental picture of where the island lies over the horizon and points his boat in the right direction until he gets there, keeping an eye on the waves and the winds and the general look of the sun and stars. Without any conscious calculation, he can keep a feel of where he should be heading while continually tacking from side to side. The Truk style of thinking is seen as primitive because the islanders are unable to articulate the rules by which they are maintaining their course.[70]

The ability to articulate and discuss publicly the rules of navigation is crucial if the activity is to be called a scientific or logical procedure. Knowledge or skills which are attributed to instinct or intuition are by definition not amenable to such public analysis. Probably the Truk islanders would not even be able to draw something so seemingly basic as a map of the islands. So, are they using a model to navigate? Does a model need to be publicly amenable to qualify as a model? Dennett describes the feed or flee response of primitive organisms as the beginning of a process of developing a point-of-view. Does this process scale up to what we call a representation of the world? The evolutionary process equips animals with the abilities to navigate and survive their environment. It is tempting to say that the adaptive nature of the evolutionary process itself is what solves the problems which the animal encounters. [71] Animals are born to fit their environment like a key fits a lock. When they find themselves in a strange environment, they often perish. In such cases their evolution-given models don't fit the strange environment. Such animals are like a robot which encounters the frame problem when its

---

[69] Ibid. p.103.
[70] Ibid. p.104.
[71] See John H. Holland "Genetic Algorithms" in *Scientific American*, July 1992.

model does not fit its altered world. The ability of human beings to survive in a range of environments is largely due to our collective ability to reshape them. Communication assisted this process and language accelerated it towards the environmental disaster which we are now witnessing, where species are becoming extinct at the greatest rate since the decline of the dinosaurs.

Artificial intelligence robots do not have the advantage of years of evolutionary adaptation to their environment, and the handy abilities that this entails. However, we haven't got 100,000 years to evolve a robot to make our snack. We need to build one with the necessary abilities already present in it. If that means building a robot with a built-in representation of its world, we need to clarify how the robot can use the model. If we agree that the insectlike abilities of Cog to navigate its world don't scale up to a model, we might be drawn to conclude that the Truk islanders are working without a model as well. That is, our definition of a model becomes a publicly discussible, logical, or even scientific construct. Does a series of contradictory, common sense statements about the world add up to a model? Lenat has programmed CYC to meditate on connections between its contradictory models. He calls it making analogies. Lenat hopes that CYC will be able to combine knowledge from one area to solve problems in another, using metaphors and analogies. This process sounds very like the one McCrone envisages happening in the human brain. These leaps across contexts, which he calls bridges between nets, are the result of thinking in analogies and metaphors.

CYC's nocturnal meditations play a role similar to that ascribed to dreams in human beings. Many psychologists and neuroscientists argue that dreams are our way of processing all the experience of our day. Some compare it to filing, some argue that it is like putting out the trash. Either way, you will agree, that dreams combine the experiences of your day in quite unexpected ways. Furthermore the process is neither rational nor commonsensical. People and places metamorphose in such a way that it is often difficult to describe who or what one was dreaming about. Males might also be female, outside might also be inside, one step might take us around the world. In our dreams we are free of the frame. When Philip K. Dick asked the question *Do Androids Dream*

of Electric Sheep? he may well have highlighted the crucial issue - can computers think illogically or creatively?

We have come full circle here from a perceived need for common sense thinking, to the idea that common sense thinking requires creativity. The above considerations concerning the evolution of human thought are speculative, and suggest that the way organisms think is largely determined by their embodiment. Brooks argues that certain analogue relations cannot be digitally modelled. McCrone argues that deductive reasoning depends on a kind of pre-logical, pre-linguistic reasoning which is already in place before human beings can use language to think. Our robot is going to need to use deductive reasoning, how do we set about installing this capacity without waiting the prerequisite 100,000 years?

The Dream of Reason

The most promising approach, and to some extent the approach taken by Dennett when tackling the problem of consciousness, lies in using advances in medicine and cognitive science to clear away the myths and the mysteries and shed light on what is possible. One thing that medicine and cognitive science have proven about the brain is that it is very adaptable. When someone has a stroke, the person can often re-learn old skills using an unaffected part of the brain. When someone gets amnesia, they do not forget how to speak, or catch a ball. The brain is layered and compartmentalised in such a way that a catastrophic event like a blow to the skull, does not knock out all functions. In evolutionary terms this isn't a surprising adaptation for a primate - apparently over 50% of tree-dwelling monkeys found dead in the wild die of a fractured skull! The brain has fantastic redundancy built in to allow for frequent knocks to the head. It is not likely that there is one way of thinking, or one way of processing sensations, and it is clear that not all our actions emanate from, or are even controlled by our brain. The brain has a remarkable capacity to rewire itself, and Dennett argues that language enables us to "virtually rewire" a lot more.

Even the so-called hard-wired motor and perception skills which human beings have are partially learned. Our nerve pathways are tuned to the outside world by stimulus from the environment for the first five years of our life. These pathways from brain to eyes, spine to muscles, only become fixed by the process of myelinisation of the pathways - a process which can take weeks or years.[72] While this process is going on, and we are acquiring motor skills and basic common sense knowledge, we are also learning language.

Language is the means whereby we plan our engagements with the world, and avoid having to learn everything from experience. It must therefore enable us to represent the world to ourselves in ways which take account of the basic common sense information about the world, as well as more advanced social and scientific information. These social and scientific models often over-ride what our intuitive and common sense "models" tell us.

[72] See *The Ape That Spoke* for an interesting account of experiments on changing the ways that kittens develop their perception of the world.

The “intuitive” thinking which characterised the Truk islanders is described as primitive because there is a tacit assumption that the scientific method (of navigation in this case) is better - that scientific thinking supersedes prior modes of thinking. What we may be finding is that scientific thinking doesn’t supersede intuitive thinking, it merely complements it. If Number 5, or Data were to be equipped to navigate the Truk islands they would use satellite positioning and charts of the area - so would we. These methods are easy to pass on and don’t require massive experience of sailing in that area. In this public sense the scientific method is the superior method, precisely because the skills are transportable. Take a Truk islander to the Orkneys or some other set of islands and they would probably die trying to navigate them. Their intuitive skills are probably specific to the Truk Islands. It may be significant that the problem with intuitive knowledge and skills is that they are often specific to an environment, and the AI problem we are facing also concerns the specialised nature of computer understanding - its brittleness.

The model of reality which science provides ranges across the whole spectrum of human experience and has powerful predictive force. It would seem to be wise to provide our robot with a good scientific model of the world, even though most people survive perfectly well with a jumble of superstitions and misinformation. The problem we noted with the scientific model was equipping our robot with the wherewithal to access information about gravity, for example, when in the vicinity of cliffs. This raises an interesting question about the level at which world-view operates in human beings. We don’t, for example, review the theory of gravitation as we accidentally knock a pencil off a table and catch it. However, if we were orbiting the Earth in a space-shuttle, Newton’s laws would be very much on our mind. If we let go of the pencil it remains floating in space. If we touch it gently it will glide through the spaceship at constant speed until it hits another object. [A material system persists in a state of motion or rest unless acted on by a force.] If we chase after it, we will need to kick-off against a wall or other fixed object. [For every action there is an equal and opposite reaction.] No doubt after a few weeks in the shuttle one would begin to move about and handle objects without having to review these laws. On the other hand, we might continue chasing pencils and crashing into bulkheads and be forcibly reminded of these laws. The internalisation of Newton’s laws involves our body and mind

working in concert to develop a skill. A tennis player has little time to review Newton's laws as the ball comes over the net, but lots of practice ensures that the player can return the ball and place it precisely. This is not instinct. The player needs to know about how balls bounce on different surfaces, the effects of top spin, and all kinds of other information about tennis balls in motion.

This knowledge is not common sense knowledge. Science regularly overturns our common sense perceptions of how the world is. It once seemed like common sense that the Sun went around the Earth, now it's common sense that the Earth goes around the Sun. It isn't instinct, or our senses, which engender this belief - as Wittgenstein once asked "How would it look if it were otherwise?" It seems reasonable to assume that if two balls of different weights are dropped from a height, the heavier ball will hit the ground first - but Galileo proved otherwise. The physics of the universe at microscopic and macroscopic levels prove to be radically counter to common sense.

If you find yourself being sucked from pole to pole through a planet that has been cored like an apple, with only a spacesuit to protect you, common sense solutions aren't very useful. In fact, you would do well to arm yourself with a reliable theory of gravity. In Gregory Benford's short story "Alphas", the hero Chansing (with his partial intelligence chip Felix), is armed with such a theory and has to make some very quick decisions to survive the ordeal.[73] In order to emerge from the south pole of Venus and not be sucked back through again, like some cosmic yo-yo, he has to calculate how much momentum light radiating onto his spacesuit imparts, and what trajectory he needs to attain in order not to be sucked back. Luckily Felix, his embedded chip, does most of the calculations.

This partnership of the "seat-of-the-pants" spaceman, Chansing, and the computer-like intelligence of Felix, is a kind of metaphor for how we look at the human mind. It has one layer which it owes to our evolution from apes and another which is enabled by language. McCrone uses the example of a falling ball to illustrate the relation of these two kinds of thinking.

[73] Gregory Benford. "Alphas" collected in *Best New SF4* edited by Gardner Dozios. London: Robinson, 1990.

> [A] commonplace misconception is that a ball falls straight to the ground if dropped while a person is walking along, whereas in fact it falls in a shallow curve because of the forward motion of the walker.[74]

If one looks at drawings of sieges from the middle-ages, cannonballs soar into the air at a fixed angle until they are above the siege-town, or opposing army, whereupon they drop directly to the ground. It seemed to the artists depicting these sieges that this was a perfectly natural account of how cannonballs behave in flight. Galileo revised this common sense view. Galileo established that cannonballs followed a curved trajectory. He thereby handed the military the knowledge to vastly improve the accuracy and destructive power of cannons. McCrone notes that fundamental errors about how the world works, such as those concerning the behaviour of objects in flight, posed no problems for early man,

> who, with the size of his brain, was already enjoying enough of an advantage over the other animals in understanding how the world worked. Of course, the greater the accuracy of perception the better, but evolution never needed absolute perfection to get brains to do a useful job.[75]

This view of human development sees language and science as partners in enabling human beings to predict and control their environment. Most animal thinking is triggered by its environment, human thinking is characterised by being able to imagine that environment some other way (to fantasise), and thereby effect change in it.

The explanatory and predictive power of scientific models has enabled human beings to transform the environment and break free of Earth's gravity and explore space. Science and technology affects lives of almost every being on this planet, and dominates the lives of most human beings. It has largely displaced religious explanations about how the universe came to be, and how it continues to be. Unfortunately, this has given rise to the notion that one is required to believe in science in much the way one was required to believe in a god, or in tenets of a faith. This confusion is epitomised in the title of the first

[74] *The Ape That Spoke*, pp 81-82.
[75] *The Ape That Spoke*, pp 81-82.

chapter of Brian Appleyard's book *Understanding the Present* - “Science works, but is it the truth?”[76].

If we are going to equip our robot with a thoroughly modern scientific model, and the linguistic capacity to use it, it would be well to examine these products. First we will look at the debate as to whether science is a ‘true’ model. We will then look at what we expect from a model. Finally, if language is the key to manipulating the model, we need to establish the relationship between language(s) and the model.

Science, Models and Language

The scientific method is the process of establishing scientific laws which, using observation and experiment, describe pervasive regularities in the world. It is widely recognised that the process has two distinct phases. The formulation of a hypothesis, and the confirmation of the hypothesis through the gathering of evidence and experimentation. Anthony Quinton describes the former phase as “inspired guessing”, and the latter as “a comparatively pedestrian and rule-governed undertaking”.[77] A more optimistic view holds that science is a continual process of discovery. Each discovery adds to a coherent scientific model which provides a systematic generalisation of the laws governing physical systems. The question is, given a different set of “inspired guesses”, might science provide a radically different model of the physical world? Is science a sociological construct?

Is Science a Sociological Construct?

One answer to this question is, yes, science is a sociological construct in the sense that science did not exist 400 years ago and can be historically documented as the product of human endeavour. There can be no debate there. But that is not the kind answer sought by those who pose the question. A growing number of scientists, sociologists and historians are arguing that the western account of the physical world is just one amongst many equally valid accounts which could have been arrived at given different socio-historic developments. Physics, they argue, might have developed in radically different

[76]Brian Appleyard. *Understanding the Present.* London: Picador, 1992.
[77] *The Fontana Dictionary of Modern Thought.* edited by Alan Bullock and Oliver Stallybrass. London: Fontana, 1977.

ways if, for example, more women had been scientists, or Christianity had not become so widespread. A number of science fiction writers have tried to imagine such alternate developments in science. Walter M. Miller explores the influence of religion on the development of the sciences in *Canticle for Leibowitz*. In A.E. van Vogt's *Quest for the Future*[78] science becomes a branch of psychology. Earthmen travel 500 years into the future to find that the physical world has been proven to be dominated by "electronic psychology". The sun, they are told, has planets orbiting around it due to its desire to achieve balance in space, and the behaviour of electrons is explained by their psychological dispositions.

Historically, from era to era, explanations for physical phenomenon have changed drastically. The "science is a sociological construct" camp are appealing to our knowledge of the history of science to suggest that the current state of science is merely a cultural and historical accident. It is as if there is an exotic universe hovering beyond our cultural blinkers which we could see if not for Newton, Einstein, and Watson and Crick. We are invited to reject the explanation that masses are attracted by gravitation, that the speed of light is constant, and that DNA carries genetic information. This suggestion, that science has constructed reality in a particular way, denies the universality of science and suggests that the nature of reality could be very different from the way science depicts it.

The real question, as Brian Goodwin notes, is "Do you believe that the object of scientific investigation is a social construct?" to which he answers,

> No. I am a realist: I believe that there is a real world that exists independently of us, although we are entangled in it, and that we can obtain knowledge about it.[79]

It is because our robot is acting in the real world that the frame problem arises. That is, it is the inability to capture reality in a model which results in our robot tripping over frame problems. The frame problem highlights the fact that knowledge is not independent of what the knowledge is used for, and seems to give weight to the argument that scientific knowledge, far from being objective, is a subjective account of the world from the point of view of 400 years of western

[78] A.E. van Vogt. *The Quest for the Future*. London: NEL, 1972.
[79] Brian Goodwin in *The Times Higher Educational Supplement*, September 30 1994.

scientists. An important point to remember about this account is that its development owes as much to accident as it does to the quest to find answers to specific questions. The telescope was developed in order that the military could see opposing armies approaching at a distance, Galileo used it to prove that the Earth moved around the sun and in doing so completely changed the course of history. The spin-off effect of the development of the telescope turned out to be more significant than the purpose it was designed for. The transportability of scientific knowledge from one sphere to another is its single greatest attribute. It transcends knowledge boundaries as surely as it transcends cultural boundaries. A midnight snack making robot doesn't need an account of the world that has been shaped by generations of snack-building robots, but it does need the kind of knowledge that can be translated into useful applications in its snack-making world. For example, the kind of knowledge it gleaned from putting blocks on top of each other may well be applicable to putting sandwiches on plates and plates on tables. Unfortunately it is not always obvious that knowledge and skills obtaining to one sphere of activity can be applied in another. One reason for this is that there can be an infinite number of ways of describing any given object in any given situation. It is only when we start asking the right questions that we start to glean useful information. We decide which are the right questions by discarding information which we regard as irrelevant. In the *Museo di Storia della Scienza* in Florence the telescope which Galileo used to make his observations of Jupiter is on display. It is part of a collection of telescopes: some long, some short, some made of metal, some of card. Some of these telescopes are covered in leather, some in marbled paper. When we start to ask the "right questions" about the nature of Galileo's telescope which of these characteristics is significant?

Objects in the world have aspects, or qualities, according to how they are observed - according to the kind of questions we ask of them. An aircraft may be, shiny, new and red. It can also be fast, expensive, and inefficient on fuel. These aspects can be noted by an observer who has the capacity to notice them, and who inquires after them. The colour of the aircraft may be obvious to any creature with colour vision, but one needs to ask how much it costs, or what kind of engine it has, to determine whether it is expensive or inefficient on fuel. We glean whatever aspects of an object are relevant to our observation. Lars-

Erik Janlert notes that the list of aspects, or qualifications, may be infinite. He writes, “How heavy the undergrowth of qualifications grows depends on how the world is categorised.”[80] One cannot possibly list all the qualities of the aircraft. One could describe it such that an engineer could build it, but such an apparently comprehensive description, wouldn’t determine if it was “elegant”. In the first instance the aircraft is categorised as a piece of engineering, in the latter as an aesthetic object. The model which we provide the engineer with is a scientific model, the description of the aircraft as an aesthetic object is a much more complex business.[81]

Objects exist in relation to each other and to the world, but their properties are determined by how we observe them - and it makes no sense to talk about objects without properties. An aircraft that is travelling at 100 knots does so whether one is observing it or not, although some observers might regard it as moving fast, others as moving slow, and others might not have the perceptual apparatus or viewpoint to detect motion at all. The way we observe the world is determined by what we require to know of it. Physicists dream up models of the physical universe in order that certain questions about properties of the physical world can be answered, and the outcome of events predicted. It is useful to describe light as an electromagnetic wave because it enables physicists to predict how light will behave in a large number of circumstances. Under other circumstances it is useful to describe light as a series of discreet "packets of light" called photons. The first of these models imagines light as travelling as a wave - spreading out like ripples in a pond - and having a presence at every point on the wave front. The second imagines light travelling like a very small billiard ball, a massless particle - being at only one place at any given time. At the present time, physicists find it necessary to work with these two models of light. Consequently, I think it is true to say that physics has been in crisis over the past 30 years because the two models seem irreconcilable and a unified model

[80] Lars-Erik Janlert. “The Frame Problem: Freedom or Stability? With Pictures We Can Have Both.” in *The Robot’s Dilemma Revisited*. edited by Kenneth M. Ford and Zenon W Pylyshyn. New Jersey: Ablex Publishing Corporation, p.42. Janlert sees the qualification problem as quite separate from the frame problem. He sees the frame problem as a problem of the form of representation of models in dynamic systems. The qualification problem applies to static systems as well as dynamic systems.

[81] See Henk Tennekes, *The Simple Science of Flight: From Insects to Jumbo Jets*, MIT Press, 1996. for some descriptions of aircraft which combine aesthetics and engineering.

is elusive. The problem of the wave/particle duality of light is at the heart of the puzzles and paradoxes which characterise quantum mechanics, and a major stumbling block on the road to that holy grail of science - the Grand Unified Theory (GUT).

John Gribbin in his book *Schrodinger's Kittens* prefers to regard the models of physics as analogies to what is happening in the physical world, and not literal descriptions.

> Indeed, it is hard to see quantum physics as anything but an analogy - the wave-particle duality being the classic example, where we struggle to 'explain' something we do not understand by using *two*, mutually exclusive, analogies which we apply to the same quantum entity.[82]

Gribbin very carefully teases out the ways in which physicists have "taken hold of the world and come up with their present description of reality."[83] His approach emphasises the historical development of models, and his own search is for a "best-buy' model. "The world may be 'like' many things - waves, or billiard balls, or whatever - without really *being* any of these things."[84]

McCrone maintains that the analogies we use to solve problems are often metaphors drawn from everyday life.

> The richer our net of knowledge about something, the better the metaphor it will make, which is why we use everyday objects to mimic the way unfamiliar or even totally abstract things are going to work. When scientists talk about electrons, for instance, they think about them in terms of waves or little balls. Electrons are not in fact much like either, but because they are beyond the limit of what we can directly sense, we have little choice but to build a second-hand picture from something familiar. Looking generally at what we consider to be abstract or rational thought, it becomes clear that we do not really understand something until we ground it in commonplace experience. A list of all the known properties of an electron remains dry words until we take a rich knowledge net about ricocheting balls or rippling waves to animate the picture. The same applies whatever the subject. The whole of human intellectual achievement rides along on the back of metaphors.[85]

This idea that logical, mathematical, and language-based thinking is driven by metaphorical and image-based thinking has many adherents. Thinkers such as

---

[82] John Gribbin. *Schrodinger's Kittens and the Search for Reality.* London: Little, Brown and Company, 1995, p.198.
[83] Ibid. p.214.
[84] Ibid. p.215.
[85] *The Ape That Spoke*. pp.107-8.

Einstein and Richard Feynmann, when pressed to describe their thought processes, often claim that they think in concrete images even when dealing with equations. Dennett recounts an amusing anecdote from *Surely You're Joking Mr. Feynmann!* where the particular mathematical problem to be solved became, in Feynmann's mind, a "hairy green ball thing". Lars-Erik Janlert believes that some combination of image perusal and inference can be used to solve the frame problem - although I don't believe it involves hairy green models of the world!

Discussing the various models of reality thrown up by theoretical physicists, John Bell asks "To what extent are these possible worlds fictions?", and continues,

> They are like literary fiction in that they are free inventions of the human mind. In theoretical physics sometimes the inventor knows from the beginning that the work is fiction, for example when it deals with a simplified world in which space has only two dimensions instead of three. More often it is not known till later, when the hypothesis has proved wrong, that fiction is involved. When being serious, when not exploring deliberately simplified models, the theoretical physicist differs from the novelist in thinking that maybe the story might be true.[86]

Successful theories of science are inevitably replaced by more successful theories, but this does not necessarily imply that science is a cultural artifact like literature, or art. The "fictions" which John Bell refers to need to be tested in reality and if they are found wanting must be abandoned no matter how elegant they seem. It is currently useful to describe electrons using both waves and balls as models. If we want to determine additional properties of electrons we may need other models - yo-yos or rubber bands perhaps. There is an infinite number of ways of looking at the physical world and an infinite number of observers - some of which have entirely different perceptual equipment. For bats, the world is a soundscape, for dogs a smellscape, and who knows how dolphins with their sophisticated sonar and acute vision, sense the world?

Gribbin never loses sight of the fact that it is theories/models/analogies which are "self-consistent and make predictions that can be tested and confirmed by experiment"[87] that are the "best-buy". The theory of natural selection has proven to be a remarkably good buy for those looking for explanations about how things work in the natural world, and is a lot more consistent than the theory that god

---

[86] J.S.Bell. *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press, 1987. Quoted in Gribbin, 1995, pp.220-221.

[87] *Schrodinger's Kittens* , p.219.

created it all in seven days. One can of course read *The Origin of Species* as a fine example of Victorian prose, or "a treatise on modes of evidence for reconstructing the past from imperfect and indirect evidence"[88], but one cannot escape its power in describing how life evolved on this planet. Richard Dawkins puts it thus,

> Darwin might have been inspired by Victorian economics when he thought of natural selection. If true, this is an interesting contribution to the history of ideas but it does not affect the primary question of whether life does, as a matter of fact, evolve by natural selection. “As a matter of fact" is not a phrase one should apologise for using.[89]

It is interesting to see the word "fact" enter into this debate, and it is significant that Richard Dawkins utters this almost taboo word. It is fashionable in many spheres of intellectual activity to regard facts as "provisional fictions", but the philosophers, sociologists, literary critics and scientists who promulgate such a notion have lost sight of the truth which Dennett mentions in passing in his essay on the frame problem - *A fact is only a fact when it is a relevant fact*. The fact that a fact is only a fact when it is a relevant fact does not make it a provisional fact. Our midnight-snack-making robot has proven that what we call facts are crucially related to what matters in the scenario we are in. Dawkins puts it in more concrete terms.

> When you take a 747 to an international convention of sociologists or literary critics, the reason you will arrive in one piece is that a lot of western-trained scientists and engineers got their sums right. If it gives you satisfaction to say that the theory of aerodynamics is a social construct that is your privilege, but why do you then entrust your air-travel plans to a Boeing rather than a magic-carpet? As I have put it before, show me a cultural relativist at 30,000 feet and I will show you a hypocrite.

The relevance of aerodynamics is very apparent when travelling in a 747 at 30,000 feet. Dawkins’ point is that western science works. If you had to choose between a western trained doctor and a medicine man to treat your diabetes, you would choose the western trained doctor. Diabetes is not caused by evil spirits, and exorcism is not a cure. The biological model developed over the last few centuries has enabled medicine to eradicate diseases, reduce infant

[88] Stephen Jay Gould on Charles Darwin's *Origin of Species* in *The Times Higher Educational Supplement*, September 30 1994.
[89] Richard Dawkins. "The Moon is not a Calabash" in *The Times Higher Educational Supplement*, 30th September 1994.

mortality, and improve nutrition. In the face of this evidence the “knowledge is a sociological construct" camp maintain that all models of reality, the Newtonian model, the relativistic model, the quantum model, even the medical model, are all convenient fictions. The debate throws into doubt the existence of everything from subatomic particles to quasars, from viruses to vitamins. Everything, so the argument goes, that science describes is an analogy for a reality we will never see and can only dimly understand through crude analogy and what our senses tell us. There is something very suspect about these arguments - all they seem to be saying is that models are just models, which is hardly a revelation! The pointlessness of the debate can be shown by examining just what comprises a model.

What is a Model?

Joseph Weizenbaum uses the example of a falling object, and the formula for acceleration due to gravity, to demonstrate the nature of computer models and how they relate to mathematical and physical models.

> The aim of a model is, of course, precisely not to reproduce reality in all its complexity. It is rather to capture in a vivid, often formal, way what is essential to understanding some aspect of its structure or behaviour. The word "essential" as used in the above sentence is enormously significant, not to say problematical. It implies, first of all, purpose. In our example, we seek to understand how the object falls, and not, say, how it reflects sunlight in its descent or how deep a hole it would dig on impact if dropped from such and such a height.[90]

What is essential to the model will be determined by the modeller according to what the modeller wants to achieve with the model. That act of choice is likely to be partly intuitive and may involve judgements based on the success or failure of past models. We will put aside for the moment problems of funding and other political and cultural issues. The modeller who designs the midnight snack scenario may choose the elements of the problem according to a whole range of expected outcomes. If the robot does not realise any of those outcomes - an

[90] Joseph Weizenbaum. *Computer Power and Human Reason*. (2nd edition), Harmondsworth: Penguin, 1984, p.149.

edible snack in an intact kitchen, for example - the experiment will probably be changed. An autonomous robot is as likely to produce a beer and chocolate sundae as a turkey sandwich, and certain experimenters might view that as creativity. Experiments are usually viewed as failures when they do not yield the expected information. A notable failure was the 1972 Viking mission to establish whether there was life on Mars. Shortcomings in the design of the experiment made its observations inconclusive. The design of an experiment pre-supposes a model of how things are. If the models omits a significant factor it will fail, because, if you will forgive the tautology, the factor is "essential". One of the problems here is that one doesn't know until the experiment fails that the element omitted was quite so relevant. Usually one does the experiment again - an option not immediately open to NASA in the Mars case.

A model with everything included is not a model, it is reality. Models are by definition incomplete - they will always lack something that the thing in the real universe possesses, even if it is only scale, or the fact that it is not actually that thing. In fact, the fewer elements there are in a model the more likely it is to be useful. Weizenbaum provides this example of a model.

$$d = at^2/2$$

This formula may be considered to be a model of an object falling towards earth. Where d = distance, a = acceleration due to gravity (32ft per second per second).

The formula says,

> The distance an object falls is equal to the acceleration due to gravity, multiplied by the time taken, squared, and divided by two.

If the object takes 4 seconds to fall, multiply 4 by itself (16) and then by acceleration due to gravity (32), and divide the result (512) by 2. The distance an object falls in 4 seconds is 256 feet.

Those who say that science is a sociological construct have 4 seconds to dodge the object, whether the model is a construct or not. Consideration of this simple model demonstrates that the debate makes no sense. Yes, the model of the object falling is constructed by human beings, but you still need to get out of the way of the falling object within 4 seconds. There are many other models which one could construct to demonstrate the above fact about the observed world.

No-one is claiming that gravity is structured by mathematics, merely that it is a very useful way of describing it.

Scientific and mathematical models are judged by their success. Success can be measured either as their ability to predict events or help clarify a problem. Human beings set these problems and judge their success or failure. Like the facts that are relevant facts in our midnight snack example, these models only discover or describe what the modellers consider relevant. One might even say "A model is only a good model if it simplifies the relevant properties." Mathematics is good at describing simple relationships. The branches of mathematics - geometry, arithmetic, algebra etc. - each focus on a different type of relationship. A model which focuses on too many properties and relationships will usually become unwieldy. Our robot needs to maintain a number of models which are simultaneously true for a number of situations and applications.

This account of models has led me to draw exactly the opposite conclusion from the "science is a sociological construct" camp. They conclude that all these models are fictions, I conclude that they can all be simultaneously true. The most extreme claim that the 'science as a sociological construct' camp can make is that the models, equations and theories which describe reality only describe a reality which human senses are capable of verifying. We have an equation for objects falling to earth because things falling to earth are part of our everyday existence. No doubt there are thousands of equations which describe circumstances which don't seem to occur in the world. The 'science is a sociological construct' camp are claiming that these could also be descriptions of a reality which we cannot perceive. They are stating what science fiction writers have been demonstrating for decades - that different species have different perceptions of the universe. Douglas Adams is fascinated by the fact that human beings assume that all creatures must perceive the world in much the same way - the way that human beings perceive it. His books constantly satirize such assumptions. His universe is full of creatures whose scale, longevity, senses, motivations and general demeanor are nothing like our own. But one doesn't need to cross the galaxy to find such bizarre creatures - the creatures which populate our planet are strange enough. Wildlife films are beginning to undermine such chauvinism, and in Adams' book *Last Chance To See*, he decided to put together his own wildlife documentary. His ruminations on the role

of our senses in prioritizing how we engage the world is focused through his encounter with a rhinoceros. The rhinoceros has nasal passages larger than its brain, it senses the world primarily by smell, its other senses, including eyesight are very poor.

> We are so used to thinking of sight, closely followed by hearing, as the chief of the senses, that we find it hard to visualise (the word itself is a giveaway) a world which declares itself primarily to the sense of smells.[91]

Our information about the world, and our assumptions about what is important in the world, are shaped by our senses. It is logical to build a robot that navigates by sonar, rather than sight, because sonar is an infinitely superior system. It just happens to be a sense human beings do not have. Navigating by sight is immensely complex and its only advantage over sonar is that it enables us to home in on coloured objects. Artificial intelligence projects are hampered by the fact that to some extent they are making square pegs for round holes. The solutions for survival which evolved the human organism are completely unsuitable for other organisms, and in the case of our robot, very unsuitable for a non-organism. When AI robots come up against a problem such as that experienced by Deep Blue, its builders redesign it to cope with a Kasparov-like strategy. Dennett comments,

> This process recapitulates the process of natural selection in some regards; it favours minimal, piecemeal, ad hoc redesign which is tantamount to a wager on the likelihood of patterns in future events.[92]

and adds in a footnote

> In one important regard, however, it is dramatically unlike the process of natural selection, since the trial, error and selection of the process is far from blind. But a case can be made that the impatient researcher does nothing more than telescope time by such foresighted interventions in the redesign system. [93]

Rodney Brooks has declared that by situating Cog in the world he seeks to emulate evolutionary processes. His manifesto paper is called *Building Brains for Bodies*, but evolution is the process of *building bodies for environments*. Furthermore, evolution is likely to come up with a number of different solutions

---

[91] Douglas Adams & Mark Carwardine. *Last Chance To See*. Heinemann: London, 1990, p.94.
[92] "Cognitive Wheels" p.145.
[93] Ibid.

for any given environment, or environmental niche. In the environment of your garden or local park you will find worms, ants, snails, spiders, birds, and lots of different kinds of beetles. Building a version of one of these creatures, a spider for example, which moves and behaves in a similar manner to a spider, is a difficult task.[94] Brooks' project is years behind schedule precisely because embodiment problems are tricky and expensive, and Cog's brain must develop with its body.

Emulating the intelligence - or thinking process - of a complex creature like a human being is difficult because there is so much about the relationship between brain and body which we do not understand. In this sense most AI projects are putting the cart before the horse (the brain before the body), hoping to leap-frog evolution and get directly to brain development. We have considered the possibility that embodiment may be crucial to the kind of intelligence we can recognise. It also seems likely that we couldn't recognise intelligence in something that wasn't embodied in a manner similar to ourselves, which didn't experience the world in a similar manner, and which didn't share some of our goals. Douglas Adams illustrates the gulf between ourselves and the rhinoceros in the following passage.

> For a great many animals smell is the chief of the senses. It tells them what is good to eat and what is not (we go by what the packet tells us and the sell-by date). It guides them toward food that isn't within line of sight (we already know where the shops are). It works at night (we turn the lights on). It tells them of the presence and state of mind of other animals (we use language). It also tells them what other animals have been in the vicinity and doing what in the last day or two (we simply don't know, unless they've left a note).[95]

Most importantly for those sneaking up on a rhinoceros, if it can't smell you, you aren't there. In Adams case, he was standing 25ft away but, despite the fact that the animal could see him(dimly), the wind direction ensured that it couldn't smell him. When the wind changed direction the three ton vegetarian suddenly ran away.

Rhinoceros physics would have developed in a very different way from that which human physics has. What does a nuclear explosion smell like? Can one

[94] See "A Cricket Robot" by Barbara Webb, *Scientific American* Dec. 1996.
[95] *Last chance to See*. pp.94-5.

smell a planet?[96]Is somebody out there developing a robot that smells its way around?

How is our robot going to make a good snack if it can't smell whether the turkey is off, or the butter rancid? What use is a snack-making robot which can't feel if the bread is fresh or the beer cold? The kind of senses or sensors which researchers build into AI robots tell us a lot about which kind of information is considered most useful - usually the kind of information communicated with light and sound. A robot that needs to make a snack should be equipped with the senses of smell and taste. It should be specialised for its environment and its activities. Evolution has generally favoured specialists. Human beings are not specialists. There are a number of theories which suggest that intelligence arises through the need to cope with different environments and changing situations. Our idea of intelligence fits very nicely the kind of beings which we happen to be - non-specialists. Clearly science has developed from that kind of intelligence, and that science has claims to being a universal science. That is, science does not claim to specialise, it aims to describe conditions in all parts of the universe as they apply to all beings.

It is conceivable that there are galaxies full of entities which have none of our senses, no vision or feeling, and, like Kurt Vonnegut's Tralfamadorians, do not exist in time as we understand it, or perhaps are not even strictly corporeal. Our description of reality will make little sense to these entities.

Looking out across the universe we can observe vast galaxies of stars, hydrogen clouds, and the remnants of the occasional super-nova - is it possible that other species see nothing like the phenomenon we see? Actually, a simple analogy shows that it is not only possible, but probable - furthermore the analogy doesn't drag us into pointless philosophical speculation on whether reality exists or not.

Chasing Rainbows: What is real and what counts as a fact?

Gribbin recounts a debate between physicists as to whether the properties of electrons, such as momentum, are real. The experimental results,

> are only telling us about the ability of electrons to respond to momentum tests, not their real momentum, just as the results of IQ measurements

[96] If cats had developed quantum theory, there would be the well-known paradox of Schrodinger's human being.

> only tell us about the ability of people to respond to IQ tests, not their real intelligence.
>
> Nick Herbert, an American Physicist, has another analogy. Bohr said that isolated material particles do not exist, but are abstractions which we only identify through their interactions with other systems - as for example, when we 'measure' the 'momentum' of an electron. This, says Herbert, is like a rainbow. A rainbow does not exist as a material object, and it appears in a different place to each observer. No two people see the same rainbow (indeed each of our two eyes 'sees' a slightly different rainbow). But it is 'real' - it can be photographed. Equally, though, it is not real unless it is being observed, or photographed. [97]

This analogy is designed to explain some peculiarities concerning the role of observation in quantum mechanics, but serves very well as a possible analogy for how reality appears on a macrocosmic scale. Those species in other galaxies who do not see hydrogen clouds and swarms of galaxies may well see something else - but this does not mean they are “fictions” or that there is nothing there.

Science fiction writers invent environments and beings to inhabit them. In this sense they perform astrophysical, geological and evolutionary roles. They build the environments and they build the creatures to fit them. One finds sentient stars and planets (e.g. Lem’s *Solaris* and Stapledon’s *Star Maker*), energy beings (*Cocoon*), shape shifters (*Star Trek: Deep Space Nine*), giant insects (*Alien),* and mind-readers, but mostly one finds bi-pedal humanoids. In general it is rare in science fiction to find a genuinely alien physics. We find planets with a different gravity from earth, more suns, colder or warmer climate, but these deviations from our norm are usually within limits dictated by the framework of popular physics.

Alien Physics: Is Science Universal?

Instances of actual alien physics in science fiction, like worlds where birds fly out of the ground,[98] or time runs backwards,[99] are rare - in fact such fiction is often classified as fantasy. A genuinely alien physics precludes any possibility of accounting for what happens in the world using current scientific theory, and the world on which it operates would strike us as bizarre. What we usually mean by alien physics is alien perceptions of physics. The philosopher of science Thomas

[97] *Schrodinger's Kittens* , p.219.
[98] “Placet is a Crazy Place” by Frederic Brown
[99] See the *Red Dwarf* Episode “Backwards”.

Kuhn, maintains that scientific communities on different worlds would arrive at descriptions of reality which do not agree with our own.

The majority of physicists, however disagree,

> They imagine that if we ever make contact with a scientific civilisation from another planet then, assuming language difficulties can be overcome, we will find that the alien civilisation shares our views about the nature of atoms, the existence of protons and neutrons, and the way the electromagnetic force works.[100]

This view asserts that modern physics has arrived at an account of the physical universe which is universal in its essentials.

H.Beam Piper's short story 'Omnilingual', illustrates this view in the tale of a group of archaeologists on Mars attempting to read an extinct Martian language. When the archaeologists discover the table of elements on the wall of a Martian university, one of them remarks,

> "That isn't just the Martian table of elements; that's *the* table of elements. It's the only one there is," Mort Trantor almost exploded. "Look, hydrogen has one proton and one electron. If it had more of either it wouldn't be hydrogen, it'd be something else. And the same with all the rest of the elements. And hydrogen on Mars is the same as hydrogen on Terra, or on Alpha Centauri, or in the next galaxy-"[101]

For these archaeologists, the table of elements is a kind of Rosetta stone which enables them to crack the Martian language. The conclusion of the story is that "physical science expresses universal facts - necessarily it is a universal language." This optimistic view of alien civilisations and alien physics makes a number of very large assumptions about alien creatures and alien cultures.

Piper's Martian civilisation died out fifty thousand years ago, but we know the Martians were oxygen breathing bi-peds, they had two sexes, lived in cities, used electricity, had universities and a highly advanced technology. 90% of science fiction depicts alien civilisations in such a way. These characteristics fit almost exactly those prescribed by C. F. Hockett in his essay 'How to Learn Martian', where he imagines how the first Martian linguist might go about the task of deciphering a Martian language.

> If there are Martians, and they are intelligent and have a language and if they do have upper respiratory and alimentary tracts shaped much like

[100] *Schrodinger's Kittens* , p.198
[101] Ibid. p.54

> our own, and ears much like ours, and, finally, if they do make use of these organs in speech communication - given all these ifs, then the procedures of Ferdinand Edward Leonard will work, and he will be able to "break" the phonetic system of the language.[102]

If one is trying to learn the language of living beings speaking a living language, the above "ifs" need to be true. The Martian archaeologists are trying to decipher a dead language and have assumed that these "ifs" are true - they have assumed that the Martians were very much like Earthlings, not just in their physical make-up, but in their social organisation.

In his novel *Babel-17,* Samuel Delany is less optimistic about the possibility of communicating with alien species. Rydra Wong, the heroine, explains the problem using the example of a race called the Ciribians. The Ciribians are a friendly, intelligent, "galaxy-hopping life- form", who because of their reproductive processes and body heat changes, have three forms of "I". Although we find out very little about these beings we are told that "Their whole culture is based on heat and changes in temperature." Because they have no word for a house or dwelling,

> You have to end up describing "....an enclosure that creates a temperature discrepancy with the environment outside of so many degrees, capable of keeping comfortable a creature with uniform body temperature of ninety-six-point-six, etc.."[103]

Conversely, they can describe a "huge solar-energy conversion plant" such that another Ciribian could build it, in nine words, "Nine very small words, too." Alien encounters are few, she explains, "Because compatibility factors for communication are incredibly low." In short, if the aliens are genuinely alien we will not have enough in common with them to learn their language. A genuinely alien species is by definition unintelligible. The view that communication through symbolic language is only possible between creatures with a similar biological and/or social make-up is borne out by the leading linguistic theories of this

Noam Chomsky has been the leading linguist in the latter half of this century. Only recently have his theories been seriously called into question. His theory of "deep grammar" and its attendant process "transformational grammar", are informed by a belief that human beings are genetically predisposed to language - that the deep grammar of language is coded in our genes

[102] Charles F. Hockett. “How to Learn Martian” The name Ferdinand Edward Leonard being a composite of the first names of linguists Saussure, Sapir, and Bloomfield.

century. Chomsky argues that human beings are genetically disposed to natural language, and Wittgenstein argues that it is agreement in judgments about "forms of life"[104] that makes communication possible. Despite the rifts which divide linguists and philosophers, no-one seriously believes that the kind of "universal translators" which operate in *Star Trek* and *Babylon-5* will ever be possible. It can be difficult to translate a text in a human language, such as French, into another human language, such as English. Much is lost in a "straight" translation, unless the text is made up of mundane sentences such as "Pass the sugar." or "Beware of the dog." Clearly if both languages have words for dog and sugar, there is a good chance of unambiguous translation. If the language-culture one is translating to has no sugar or dogs, other strategies are required. If this other language-culture has no nouns or verbs, doesn't recognise objects as separate entities, or finds actions obscene, one must start from basics, and that means looking for commonalities in culture.

Chomsky would rule out the universal translator on the grounds that it is our genetic make-up which determines universal human grammar, and to some extent determines how we apprehend the world. Creatures which do not share genetic make-ups cannot arrive at similar accounts of how the world is.

The Wittgensteinian view is that creatures who share a form of life can communicate. The Wittgensteinian view doesn't exclude the universal translator, but makes it very unlikely.

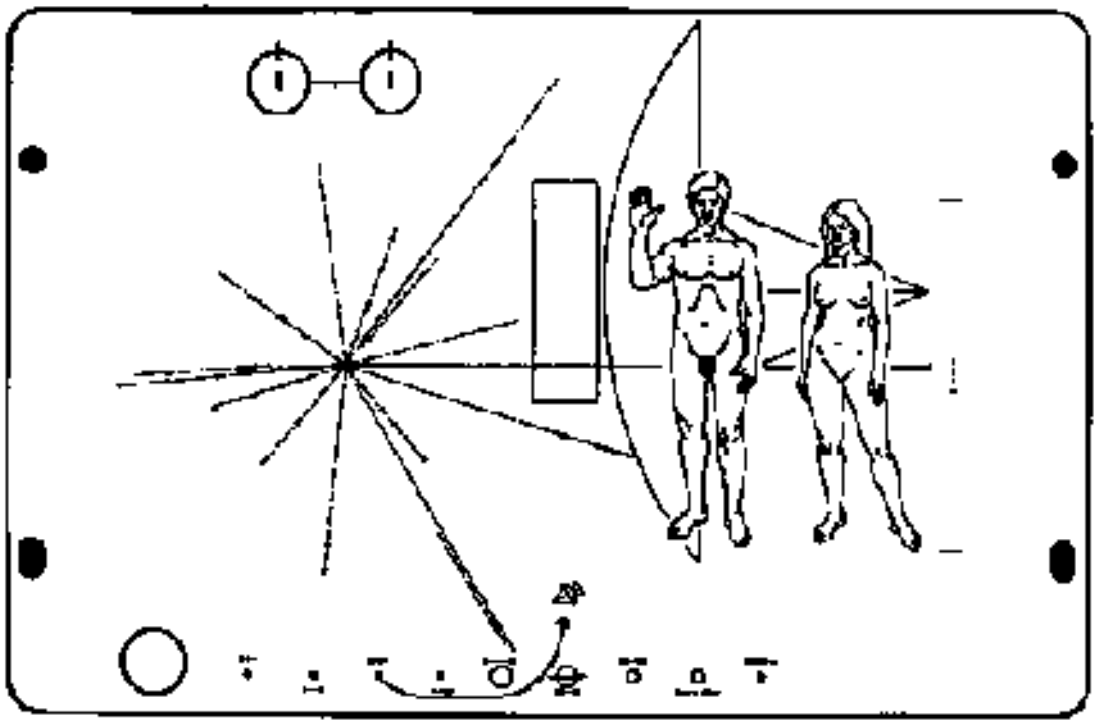
We have moved from considering whether species on other planets would be led to a different account of physics, to considering how language maps onto that account. If the Ciribians had a "table of elements", that is, if they recognised that there were different molecular structures, how would they differentiate them? Instead of classifying them according to the number of electrons and protons each had, they would be more likely to classify them according to how they responded when heated. If the Martian table of elements, in the Piper story, had been classified in such a way, the archeologists would have been unable to crack the number code, and unable to decipher the language.

---

[103] Samuel Delany. *Babel-17*. rpt. London: Sphere, 1966, p.112

[104] See *Philosophical Investigations*. By "forms of life", Wittgenstein means activities and ways of living in a society.

Do the languages of these cultures, and human cultures, reflect the biases of their perceptual apparatus? Or is it merely that one's view of the world is determined by ones perceptual abilities, and language merely reflects this? Is there something in the structure of a language which reflects the structure of the world it describes and is part of? It seems logical that language should exhibit the "multiplicity" of the world it is used to describe.[105] It should map onto that world in a way that reflects the elements in it and their relationships. It seems like common sense that in our world there is matter, relations between matter, actions, and properties - these things roughly corresponding to nouns, conjunctions, verbs, adjectives and adverbs. All this is wrong. Language does not relate to the world in anything like this manner. Language is just one way of representing the world - other modes are drawing, painting, sculpture, photography, film, sound recording, maybe even music. There is no necessary connection between a mode of representing and the thing represented. The way we choose to represent a model of the changing world to a robot is crucial - but is it necessary that the structure of the language somehow reflects the structure of its world? Consider how the drawing below tells us that some of the marks represent a man and a woman?



[105] The "picture theory of language" put forward in Wittgenstein's *Tractatus* embodies the idea that languages are projections of the elements of the world together with their relationships.

Men and women do not have black lines around them, they aren't flat and white, they aren't two inches high, and they don't float around in mid-air. There is very little in the drawing which reflects what we know about men and women, who are three dimensional; variously coloured, hairy, smelly etc. The drawing is etched on a plaque on the side of the Pioneer series of space probes and assumes that a species which encounters the craft can read line drawings, understand mathematics, and identify sexes, amongst other things.[106] Reading a drawing, painting or photograph, requires training in recognising how images represent. However, the rules that govern how images represent, and how languages represent, are not something intrinsic to the world, the image, the language, or to the system.

One might speculate that at some time in the distant past, one of our ancestors drew a line in the sand to represent a river, but it could also have represented a mountain ridge, a goat track, or it might have merely been a line in the sand. There is nothing intrinsic to the line which makes it a river, it can only represent a river if a whole range of other things are the case e.g. there are rivers, rivers are significant, rivers look like lines to you. Rivers do not look like lines to ants, lizards and ground dwelling creatures in general.

My speculation on the nature of alien cognition is not a cry for an alien AI program, it is designed to highlight the arbitrary nature of representational systems. Representational systems do not relate to the world in a way that is either intrinsic or natural. In the above drawing it is the silhouette of the man and woman which enables us to identify them as such. Our familiarity with shadows and lighting effects makes it possible for us to read drawings and photographs. Any given mode of representation is going to capitalise on how we use at least one of our senses. The representation will simplify relations in the world by selecting relevant aspects to represent. In order to survive in a changing environment, an agent needs to:-

recognise - plan - act

He later repudiated this theory when he wrote *Philosophical Investigations*.

[106] In the December 1994 issue of *Omni* there is an article on the Richard Hoagland, who together with Eric Burgess, designed the plaque.

Primitive organisms omit the "plan" part. In order to plan, the agent must represent the relevant aspect of the situation to itself so that it can "look before it leaps". This mode of representation might be as wild as Feynmann's "green hairy ball", or as apparently logical as the equation for acceleration due to gravity. Recognising what is the relevant aspect of the situation to model is the hard part. We saw with Deep Blue that Kasparov was easily able to outwit it with a little misdirection. Deep Blue has an exhaustive database of similar scenarios but was fooled by just one misplaced piece. Dennett imagines that a robot loaded with stereotypical midnight snack scenarios, including subroutines for spreading mayonnaise, pouring beer, etc. might also misanalyse some uncooperative element of the situation and be led unwittingly into misadventure.

> The shortcuts and cheap methods provided by a reliance on stereotypes are evident enough in human ways of thought, but it is also evident that we have a deeper understanding to fall back on when our shortcuts don't avail, and building some measure of this deeper understanding into a system appears to be a necessary condition of getting it to learn swiftly and gracefully.[107]

This echoes Kasparov's comment that Deep Blue doesn't really understand chess, it merely aims for materialistic goals.

I'm not sure if Deep Blue's difficulty defeating a grand master is comparable with a robot's inability to keep track of which block is on top of which other block. A limited number of rules govern Deep Blue's world, these are the constitutive rules of chess. Deep Blue's pattern recognition is based on comparing each position with its vast database, and even its designers admit that it still can't identify patterns in the game the way a human chess player can.[108] Furthermore, as Kasparov points out, it has not grasped the unwritten rules of strategy. *Scientific American*'s anonymous commentator on the second Kasparov/Deep Blue match notes that Kasparov took a while to decide which pawn to use when capturing a bishop. Using the f-pawn would have split his pawns into two masses and made them harder to defend. Secondly, it would have caused his king to have to castle short into the region where Deep Blue was concentrating its attack. Kasparov captured with the alternate pawn, in a move which the commentator describes as a "quieter, more solid

[107] "Cognitive Wheels," p.145.
[108] Interview for *Scientific American* at http://www.ibm.chess.com

capture".[109]The game resulted in a draw, but Deep Blue won the match. When Deep Blue was deciding its moves you can be sure that the above kind analysis did not enter into its calculations. In this latest contest, having no strategy did not prevent Deep Blue from winning, but in the real world having no strategy is often fatal.

The rules which govern the behaviour of things in the real world need to be discovered (and science is trying hard to discover them), but even these rules might turn out to be like chess's constitutive rules, they might provide no clues as to how to engage the world strategically.

The power of mathematics to describe physical phenomenon fosters the illusion that the observable universe is in some sense reducible to equations. The apparently universal nature of mathematics seems to support such an idea. Weizenbaum argues that this common error of assuming that the model contains all the properties of the thing modelled is responsible for many of the failures in AI research. If we could build a computer model of the human brain which is describable in strictly mathematical terms this does not imply, Weizenbaum argues, that the language our nervous system uses must be the language of our mathematics. He goes on to quote John von Neumann

> "When we talk mathematics, we may be discussing a *secondary* language, built on the *primary* language truly used by the nervous system. Thus the outward forms of *our* mathematics are not absolutely relevant from the point of view of evaluating what the mathematical or logical language *truly* used by the central nervous system is"[110]

It is interesting that this computer pioneer was not under the illusion that the logical structure of his computer reflected that of the human brain.

The idea that thinking is patterned by some kind of mental language has been given a lot of credence in recent years. Von Neumann rejects the idea that a particular mathematical language is used by the human nervous system, but suggests that some kind of language is used. Chomsky believes that structures operating at a genetic level and predispose us to language acquisition. According to Chomsky, our models of the world are largely delimited by what he

[109] Scientific American Commentary on Kasparov vs Deep Blue on May 4th 1997. http://www.sciam.com/explorations/042197chess/050597/chesscom.html
[110] *Computer Power and Human Reason*, p.150.

describes as “deep grammar” - a patterning of language not at the level of English or Chinese, but at a deeper, genetic level.[111]

If our engagements with the world require a model (or frame), however tacit or incomprehensible, and that model is in any way structured by language, it would be useful to know which parameters are determined by our genes. The spectre which looms here is that of being programmed. If we are genetically coded to see the world in a particular way, then we damned well want to know about it! This spectre in one form or another has dogged western philosophy since the time of Descartes. It raises questions about free will, the limits of knowledge and the role of language. It also tantalises AI researchers with the promise that language holds the key to enabling their creations to navigate the world.

3. Knowledge and Language

In *Problems of Knowledge and Freedom* Chomsky argues that a kind of genetic syntactical structuring determines the nature and degree of freedom which the human mind has in apprehending the world. He maintains that a series of structure-dependent operations on various sentences can show that there are underlying structures in all human beings, which predispose us to structure our perceptions of the world in a particular way. He writes,

> Thus in an important sense the rules are 'structure dependent and only structure dependent,' Technically, they are rules that apply to abstract labelled bracketing of sentences (abstract, in that it is not physically indicated), not to systems of grammatical or semantic relations. Again, there is no *a priori* necessity for this to be true. These characteristics, if true, are empirical facts. It is reasonable to suppose that they are *a priori* for the organism, in that they define for him, what counts as a human language, and determine the general character of his acquired knowledge of language. [112]

What Chomsky is saying is that grammatical rules are often structured in ways which are illogical and meaning independent. He proposes that the genetic structures which predispose us to language are similarly non-meaning related. They are not structures which can be said to relate to how the world is. Nevertheless, they determine how and what we can know of the world. It is as if we see the world through a pair of arbitrarily coloured glasses.

---

[111] Noam Chomsky, *Problems of Knowledge and Freedom*. London: Fontana, 1972.

Chomsky's approach is novel in that he cites the often illogical and meaning-independent quality of many of the rules which he identifies, as an indication that they are the manifestation of something deep-seated rather than culturally imposed. These principles are, he imagines, laid down in our biological make-up.

> Perhaps this means that the innate schematism that the child brings to bear in language learning is unique to language. If so, the neurologist faces the problem of discovering the mechanisms that determine this schematism, and the biologist the problem of explaining how these developed in the course of human evolution.[113]

It ought to be noted here that Chomsky does not state that the capacity for language evolved because language gave human beings an evolutionary advantage. Chomsky remains agnostic when it comes to the adaptionist nature of what is sometimes called "the language organ."[114] The capacity for language might be a side-effect of another adaptation. Having set the agenda for late 20th century linguistics, Chomsky's agnosticism has surprised many, and fuelled a lively debate about the adaptionist nature of our language capacity.

Chomsky suggests that a biological mechanism which predisposes us to language may impose "absolute limits on what can be known". In his 1988 Managua Lectures the following assertion sets the agenda for discussion.

> A person who speaks a language has developed a certain system of knowledge, represented somehow in the mind and, ultimately, in the brain in some physical configuration.[115]

He sees it as the job of linguist-psychologists to pursue the issues that arise from his assertion, and set the stage for further enquiry by brain scientists into the physical mechanisms which determine how our minds work. He see this project as a "step towards assimilating psychology and linguistics within the physical sciences."[116] Chomsky and Dennett seem to have something in common here in that they suspect that there are innate mechanisms which determine what we know and can know, and that scientific enquiry can throw light on how they work.

---

[112] *Problems of Knowledge and Freedom*. p.23.

[113] Ibid., p. 44.

[114] A recent letter to *Nature Genetics (*Volume 18, Number 2 - February 1998*)* from a British team working at the Welcome Trust Centre for Human Genetics in Oxford suggests that they are close to discovering which genes which control the development of language. Their work involves the study of a large 3-generation family who exhibit similar speech and language disorders. They have designated the gene they are seeking SPCH1.

[115]Noam Chomsky. *Language and Problems of Knowledge: The Managua Lectures*. MIT Press, 1988, p.3.

Dennett tentatively suggests that there may be a number of things which we are "born knowing", and these things may well be the crucial frame information that our robot needs. He suggests that knowing that once the beer is in the glass it can no longer be in the bottle may be a piece of the frame we are born with, but he is hesitant to call such information knowledge. Both Chomsky and Dennett suggest the possibility that we are biologically pre-disposed to certain kinds of "knowledge", above and beyond the predispositions which our senses determine.[117] We are born with the frame to some extent built into us. This opens the door to AI researchers who argue that when they program a robot with common-sense information they are not doing its thinking for it, merely duplicating the condition of nature. What we call intelligent beings start with the commonsense information already installed.

The hypothesis that our perceptions of the world are in some way programmed is the subject of Dick's "The Electric Ant". Instead of waking up, as Gregor Samsa did in Kafka's *Metamorphosis*, transformed into a giant beetle, Garson Poole wakes up and is informed,

> You're a successful man, Mr. Poole. But, Mr. Poole, you're not a man. You're an electric ant.[118]

An electric ant, it transpires, is a humanoid organic robot, programmed with the delusion that it is human. When Garson Poole realises he is a machine he begins to speculate on the nature of free will.

> Shall I go to the office? he asked himself. If so, why? If not, why? Choose one. Christ, he thought, it undermines you, knowing this. I'm a freak, he realised. An inanimate object mimicking an animate one. But he felt alive. Yet .... he felt differently, now. About himself. Hence about everyone, especially Danceman and Sarah, everyone at Tri-Plan.[119]

Garson Poole immediately sets about a series of experiments designed to discover how he knows things. Like Chomsky, he is interested in the mechanism whereby he perceives reality. Unlike Chomsky, he finds it in a "punched tape roll"

---

[116] Ibid. p.5.

[117] It ought to be noted here that Dennett does not regard the debate about whether such knowledge is innate or learnt to be a crucial issue for Cognitive Science. Dennett attacks Chomsky for his agnosticism concerning the evolutionary, adaptionist nature of the development of what he describes as "the language organ".(Dennett, 1996).

[118] Philip K.Dick. "The Electric Ant." *The Magazine of Fantasy and Science Fiction*, 1969. Rpt. in *Machines That Think*, p.497.

[119] Ibid., p. 499.

above his heart mechanism. This is his "reality-supply construct". The computer which Poole hires to diagnose his problem explains,

> "All sense stimuli received by your central neurological system emanate from that unit and tampering with it would be risky if not terminal." It added, "You appear to have no programming circuit"[120]

The idea that he is being controlled by a "reality tape" is so repugnant to Poole that the first question he asks is,

> Do I want to interfere with the reality tape? And if so, why? Because, he thought, if I control that, I control reality. At least so far as I'm concerned. My subjective reality...but that's all there is. Objective reality is a synthetic construct, dealing with a hypothetical universalization of a multitude of subjective realities.[121]

Chomsky notes a particular difficulty that confronts brain scientists in investigating the physical mechanisms involved in representation, acquisition and use of knowledge - for ethical reasons they can't experiment on human brains. In effect, they are forbidden from interfering with people's "reality tapes".

> We do not permit researchers to implant electrodes in the human brain to investigate its internal operations or to remove parts of the brain surgically to determine what the effects would be, as is done routinely in the case of non-human subjects. Researchers are restricted to "nature's experiments": injury, disease and so on. To attempt to discover brain mechanisms under these conditions is extremely difficult.[122]

Another problem is that only human beings seem to possess the language faculty. Study of the brain mechanisms of other animals throws little light on this crucial faculty of the mind/brain of human beings.

Garson Poole has no such restrictions. He is able to experiment on his brain and observe how reality changes as he alters parts of its mechanism. Poole's conclusion that there is no such thing as objective reality, merely a multitude of subjective realities which sometimes coincide due to the universal nature of perceptual apparatus. This conclusion is similar to Chomsky's, who also argues that the "principles of mental organisation" enable us to achieve consensus about the nature of reality.

> The principles of mind provide the scope as well as the limits of human creativity. Without such principles, scientific understanding and creative acts would not be possible. If all hypotheses are initially on a par, then no

[120] Ibid., p. 501.
[121] Ibid., pp. 501-502.
[122] *Language and Problems of Knowledge*. p.136.

> scientific understanding can possibly be achieved, since there will be no way to select among the vast array of theories compatible with our limited evidence and, by hypothesis, equally accessible to the mind.[123]

This view of the mind as a kind of valve which controls the influx of an anarchic reality, provides a picture which leads us to doubt the veracity of our senses and leads to questions about the real nature of reality. The picture is misleading because it pretends that it might be possible to apprehend reality directly, i.e. without these 'censoring' devices. This is exactly what Garson Poole attempts to do. By punching holes in his reality tape, by inserting blank bits, and finally by cutting it all together, he makes various aspects of his reality appear and disappear until he experiences "absolute and ultimate reality", and "dies".

Garson Poole is a robot which attempts to control its reality by altering its programming. It finds that it is programmed with all the "frame" information that AI researchers find so difficult to generate, and systematically slices away at it until reality disappears: effectively, it removes the frame. The idea that one can directly apprehend reality without the censoring devices of the senses and the mind is nonsensical. It is like wanting to experience the weather - but not any particular kind of weather.

"The Electric Ant" is an exploration of the frame problem in reverse, and strongly parallels the efforts of deconstructionists such as Roland Barthes and Jacques Derrida to slice away at the underlying structures of language and narrative in order to escape the ideological biases built into language. Like Garson Poole, they find the idea of being programmed repugnant!

Dick's protagonists are a reflective bunch, and his androids and robots are no exception. The replicants in *Do Androids Dream of Electric Sheep?* also have the ability to reflect on the nature of their consciousness - How are they programmed? Have they got free will? In what way are they different from human beings? These issues arise with Dick's androids as they do with most robots who have mastered the trick of talking to themselves.

Consider the following line of reasoning:-

The possession of language enables robots to reflect.

---

[123] *Problems of Knowledge and Freedom*, p.45.

If they can reflect on their own position they are conscious.

If they are conscious they have free will.

If they have free will they are free of their programming.

If our robot has language it will be a free agent. Number 5 and Data have language and are free agents, yet they still occasionally trip over the frame problem. Neither being a free agent or having language guarantees a grasp of the frame. The aliens in *3rd Rock From the Sun* have language, and their lives on Earth are just one big frame problem. Mr. Bean has enormous frame problems and he is a human being (we assume).
Chomsky believes that the structure of "the language organ" is the basis for our system of knowledge. The language faculty therefore determines the scope and limits of the human mind. He also notes that paradoxically, we can never know what these limits are because it is language which enables us to think freely. Weizenbaum's account of Chomsky's project is particularly lucid.

> Chomsky's most profoundly significant working hypothesis is that man's genetic endowment gives him a set of highly specialized abilities and imposes on him a corresponding set of restrictions which, taken together, determine the number and kinds of degrees of freedom that govern and delimit all human language development. [124]

This is not very useful for us when we build our robot, because Chomsky's account of language does not contain an explicit model of the world. Other philosophers and linguists have not been so reticent, and have proposed that grammar at a lexical level - not at some deep biological level - structures how we see the world. For these thinkers language sets the agenda, and traps speakers in a linguistic prison.

[124] Weizenbaum (1984). (pp.136-137)

"The question is," said Alice, "whether you can make words mean so many different things."
"The question is," said Humpty Dumpty, "Which is to be the master - that's all." (Through the Looking Glass Ch- 6).

Distrusting Language

If our robot possessed language it would be able to understand English language commands, tells us about its day, and presumably ask us questions about its programming. Would a robot which had learned to reflect, suddenly begin to sulk about the nature of its existence? I noted earlier that a robot capable of getting bored would easily break out of a logical loop. Dennett characterises consciousness as a kind of free gift that came with language. Language is thus a two-edged blade - it enables us to plan our next action, but it also enables us to plan not to act at all. Existentialists like Sartre and Camus were particularly good at justifying plans to do nothing - chiefly on the grounds that there was no point in doing anything because eventually we will die and be eaten by worms. This strand of French nihilism took a strange turn in the 1960 when literary critics began to suspect that language was somehow preventing them from breaking out of the existentialist dilemma which it had paradoxically caused.

In his Inaugural Lecture to College de France in 1977, Roland Barthes expresses his distrust of language, and his suspicion that he is being railroaded into a particular way of thinking by language,.

> Jakobson has shown that a speech-system is defined less by what it permits us to say than by what it compels us to say. In French (I shall take obvious examples) I am obliged to posit myself first as subject before stating the action which will henceforth be no more than my attribute: what I do is merely the consequence and consecution of what I am. In the same way I must always choose between masculine and feminine, for the neuter and the dual are forbidden me.[125]

Throughout his lecture, Barthes uses slogans like "Language is legislation, speech is its code." and "Language.....is quite simply fascist." Barthes wishes to "abjure" the power of language, to "cheat" it, and enjoy speech "outside the boundaries of power" through literature. Barthes describes literature as "a permanent revolution of language" and sees salvation from the tyranny of

language in the "play of words." The term "play" echoes the title of Derrida's post-structuralist manifesto, "Structure, Sign and Play in the Discourse of the Human Sciences", and is a key term in the deconstructionist project. Barthes emphasises the "play" of language, in opposition to language used as an instrument to convey a message. Derrida describes play as "the disruption of presence", and language as a "field of infinite substitutions". The play of language is,

> the joyous affirmation of the play of the world and of the innocence of becoming, the affirmation of a world of signs without fault, without truth, and without origin which is offered as an active interpretation.[126]

Both of these writers celebrate an essentially non-functional aspect of language. Derrida notes that words are always metaphors and substitutes. His complaint is that the words come laden with meanings which derive from metaphysics and science, and that he, Derrida, wishes to divest these words of such accretions and strike out into ideologically pure territory. He, like Barthes, realises the impossibility of his project (to get outside language) and so sets out to deconstruct language by revealing the contradictions and assumptions which are inherent in terms and their associated concepts. Derrida revels in the fact that the task of deconstruction is impossible because the language which deconstructionists use undermines their project at every turn. He writes,

> we cannot utter a single destructive proposition which has not already slipped into the form, the logic, and implicit postulations of precisely what it seeks to contest.[127]

The deconstructionist's radical distrust of language is rooted in a belief that language operates like a pair of blinkers making us see the world in a particular way, and directing our actions accordingly. They believe that taking language to pieces might give them a glimpse of the reality beyond the blinkers - or more properly an alternative false reality. Chomsky also believes that language predisposes human beings to a particular way of apprehending

---

[125] Roland Barthes. "Inaugural Lecture to College de France" (1977) reprinted in *A Barthes Reader.*

[126] Jacques Derrida, "Structure, Sign and Play, in the Discourse of the Human Sciences" in *Modern Literary Theory: A Reader* (2nd Ed) edited by Philip Rice and Patricia Waugh, London: Edward Arnold, 1992, pp. 149-165. This paper was originally delivered at Johns Hopkins University in 1966.

[127] Ibid.

the world. Neither Chomsky nor Derrida are claiming that language programmes their every action, but they are claiming that frame information about how things relate to each other in the world, and the ability to see oneself as an agent in the world, are forced upon us by language. In their view, language shapes our thinking, and determines our world-view, but they are reluctant to be specific about where programming ends and freedom begins. If language came equipped with an obligatory world-view, our problems would be solved. When we equipped our robot with language, we would also be equipping it with a model. The theories of Roman Jacobson, Benjamin Lee Whorf and George Orwell support this view.

One wonders...what makes the notion of linguistic relativity so fascinating even to the non-specialist. Perhaps it is the suggestion that all one's life one has been tricked, all unaware, by the structure of language, into a certain way of perceiving reality, with the implication that awareness of this trickery will enable one to see the world with fresh insight. John B. Carroll[128]

Linguistic Relativity, Newspeak, and Babel-17

In Orwell's Nineteen Eighty-Four the ruling party, Ingsoc, design a language called Newspeak to restrict and control thought. The purpose of Newspeak, according to the anonymous author of the appendix of Nineteen Eighty-Four was

> ...not only to provide a medium of expression for the world view and mental habits proper to the devotees of Ingsoc, but to make all other modes of thought impossible.
> A person growing up with Newspeak as his sole language would no more know that 'equal' had once had the secondary meaning of 'politically equal', or that free had once meant 'intellectually free', than for instance, a person who had never heard of chess would be aware of the secondary meanings attached to 'queen' and 'rook'. There would be many crimes and errors which it would be beyond his power to commit, simply because they were nameless and therefore unimaginable.[129]

Newspeak is deliberately constructed to serve the political ends of Ingsoc, and achieves its effect chiefly through reducing the vocabulary, and twisting the meanings of the vocabulary that remains. The concept of Newspeak relies on the assumption that one cannot think something if one does not have a word for it. For example, if one does not know the word "jealousy" then one cannot feel jealous.[130]

Roman Jakobson forwards a view of language which is almost the exact opposite of that represented by Newspeak. He suggests that we regard language as a storehouse of codes,[131] which is a little like saying that to express an idea, or describe something, we simply visit our mental supermarket of words and pick the ones for the job. This view, that language gives expression to pre-

[128] John B. Carroll in his Introduction to *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf*. The M.I.T. Press, 1956
[129] George Orwell, *Nineteen Eighty-Four,* Harmondsworth: Penguin, 1949
[130] See B.F.Skinner, *Walden Two* , New York, 1948, for an examination of this idea.
[131] See *The Fundamentals of Language* by Roman Jakobson and Morris Halle , The Hague: Mouton, 1956.

linguistic thinking, has many adherents. Curiously, Orwell also espouses this view in his essay "Politics and the English Language".[132] He talks of "letting the meaning choose the word", and suggests putting off "using words for as long as possible" in order to "get one's meaning as clear as one can through pictures and sensations".[133]In *Consciousness Explained* Dennett attacks this view and argues that to some extent speaking is thinking. He argues that if the language we speak is the expression of our thoughts, then there needs to be a language of thought (which he calls “mentalese) with which we do our thinking. He argues against the existence of “mentalese” and emphasises that much of what we think is only made possible by language.

These are the two extremes of language theory:-

1. language voices pre-linguistic experience, it is the expression of thought.

2. language determines what we can think, it is the tracks on which thought runs.

The latter view, embodied in Newspeak, implies that the world-view of native speakers of a language is determined by the grammar and vocabulary of that language. This theory of language is known as linguistic relativity and is a view posited by the linguist Benjamin Lee Whorf in the 1940’s. His research posed the question "Does our native language shape the way that we look at the world?" and his theories answer this question in the affirmative. Whorf's theory asserts that the native grammar of a linguistic community pre-disposes them to a particular world-view. His famous example is that of the Hopi indians, whose language apparently has no tenses, and no word for time. Whorf states,

> The Hopi language is seen to contain no words, grammatical forms, constructions or expressions, that refer directly to what we call "time," or to past, present or future, or to enduring or lasting, or to motion as kinematic rather than dynamic[134]

and concludes that the "Hopi who knows only the Hopi language",

> has no notion of TIME as a smooth flowing continuum in which everything in the universe proceeds at an equal rate, out of a future, through a present, into a past; or, in which, to reverse the picture, the observer is being carried in the stream of duration continuously away from a past and into a future.....

---

[132] George Orwell. "Politics and the English Language"

[133] Ibid.

[134] Benjamin Lee Whorf. "An American Indian Model of the Universe" in *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf* edited by John B. Carroll. MIT Press: Massachusetts, 1956, p.57.

> At the same time, the Hopi language is capable of accounting for and describing correctly, in a pragmatic and operational sense, all observable phenomena of the universe.....
> Thus, the Hopi language and culture conceals a METAPHYSICS, such as our so-called naive view of space and time does, or as the relativity theory does; yet it is a different metaphysics from either.[135]

Whorf's theory could be used to support Barthes' complaint that language pre-disposes him to construct himself, and the world, in a particular way. If our native grammar shapes our view of the universe, and how we describe it, it seems logical that different language groups would tend toward different world-views. Orwell's Newspeak restricts the thought of its speakers chiefly through reducing Newspeak's vocabulary to those words acceptable to the party. The Hopi example suggests that more radical restrictions on thinking can be effected through grammar, although Whorf's conclusion that the Hopi had no concept of time seems rather far-fetched.[136] It transpires that the Hopi have three formal tenses which Hockett calls the nomic, the reportive and the expective.[137] The nomic is used in assertions of something unchanging - the height of a mountain or the colour of the sky. The reportive belongs to historical assertions - events about which we have information. The expective is appropriate to the realm of the intermediate, the anticipated or the planned.

If linguistic relativity were true, Hopi would be the perfect language with which to programme our robot. The grammar of Hopi automatically breaks up the world into things that don't change, things that we have learnt, and things that may happen. If our robot's language could automatically do this, our robot wouldn't spend all its time worrying about non-effects - like the mayonnaise changing colour - and free up its thinking for things that do change. Janlert suggests that our "operative metaphysics" is partially reflected in natural languages. What he means by metaphysics here concerns questions of the order,

> What are the fundamental entities? What are the fundamental presuppositions? What basic categories are there? What basic principles?[138]

---

[135] Ibid. p.58

[136] In a world without time there would be no frame problem. Time, on the other hand, is an abstract term which we use when we need to talk about movement and change. If nothing in the universe moved - not even an electron - we wouldn't need the concept of time.

[137] Charles Hockett. "Information, Entropy, and the Epistemology of History" in *The View From Language.* Athens: Univ. of Georgia Press, 1977.

[138] "Modelling Change" p.31

More particularly he asserts that the view of reality as a series of situations following each other (like a series of snapshot or film frames) is not reflected in our common sense view of the world or in natural language.

> natural language is constructed on the assumption that we are dealing with things that are extended in time, but undergo changes[139]

It is tempting to think that language, or at least a language, could provide the mode of representation of the frame relationships - that its grammar would make relations in the world implicit. Samuel Delany explores this idea in his novel *Babel-17* where he invents an artificial language called Babel-17. We are told that the manner in which Babel-17 is constructed makes the perception of certain relationships in the observed world unavoidable. Babel-17 programs whoever learns it to sabotage the war-effort of the alliance - that program is part of its grammar. In order to make the person an efficient saboteur, the language is an exact analytical language which "almost assures you technical mastery of any situation you look at."[140] When the heroine, Rydra, thinks in Babel-17, she is able to break free of a complex restraining device, and analyse the invaders' defence formations, simply by looking. It is the grammar of the language which makes her perceive the complex relationships. Delany is suggesting here that just possessing a language can define certain relationships in the world.[141]
In learning Babel-17, Rydra unwittingly becomes a saboteur and even sabotages her own spaceship. She explains that the reason for this is that Babel-17 has no "I".

> The lack of an "I" precludes any self-critical process. In fact it cuts out any awareness of the symbolic process at all - which is the way we distinguish between reality and our expression of reality.[142]

Rydra argues that because Babel-17 has no “I” it acts like a computer language. The person who knows it is programmed to react in a certain way to certain stimuli. Because the person is thinking in Babel-17 and this language has no "symbolic process", the word <u>is</u> the thing. As Rydra says, "the lack of an "I"

---

[139] Lars Eric Janlert in an email to the author 1997.
[140] This comment recalls Wittgenstein's statement, "To master a language is to master a technique."
[141] For an interesting discussion of this idea and Babel-17 in general, see “Could Anyone Here Speak Babel-17” by William M. Shulyer, Jr., in *Philosophers Look at Science Fiction* edited by Nicholas D. Smith, Chicago: Nelson-Hall, 1982.

blinds you to the fact that although it's a highly useful way to look at things it isn't the only way."[143] Rydra has defined the difference between an observer-centred, and object-centred representation. Janlert puts it like this,

> in an observer-centered representation of the world, when the agent takes a step forward, the whole world changes (except the agent itself), whereas in an object-centered view the world stays the same, only the agent's position has changed.[144]

A language without an "I" seems to put the agent in an object-centred world, a world in which it is just part of the situation. It cannot reflect on its position in the world because it hasn't got a concept of self. It is the world that is stable and the agent that is changing. In an observer-centred representation, the agent is stable in an ever changing world - this is at least the beginnings of a sense of self. Descartes ruminations on how we can know that we or anything else exists lead him to this simple affirmation of self - "I think, therefore I am." In John Carpenter's 1974 film *Dark Star,*[145]a computer bomb follows a similar line of reasoning when a series of malfunctions on a star-ship cause one of its thermonuclear bombs to prime itself to detonate whilst still attached to the ship. The ship's main computer "Mother", manages to get it to return to the bomb bay twice, but bomb #20 has only one destiny - to explode. Efforts to convince the bomb that its orders are faulty fail, so Commander Doolittle tries to teach it "a little phenomenology".

Doolittle:	Hello, Bomb? Are you with me?
Bomb #20:	Of course.
Doolittle:	Are you willing to entertain a few concepts?
Bomb #20:	I am always receptive to suggestions.
Doolittle:	Fine. Think about this then. How do you know you exist?
Bomb #20:	Well, of course I exist.
Doolittle:	But how do you know you exist?
Bomb #20:	It is intuitively obvious.
Doolittle:	Intuition is no proof. What concrete evidence do you have that you exist?
Bomb #20:	Hmmmm.....well.....I think, therefore I am.
Doolittle:	That's good. That's very good. But how do you know that anything else exists?

---

[142] *Babel-17, p.154*
[143] *Babel-17, p.155*
[144] "The Frame Problem" pp.39-40.
[145] *Dark Star* directed by John Carpenter and written by John Carpenter and Dan O'Bannon, 1974.

Bomb #20: My sensory apparatus reveals it to me. This is fun!
Doolittle: Now, listen, listen. Here's the big question. How do you know that the evidence your sensory apparatus reveals to you is correct? What I'm getting at is this. The only experience that is directly available to you is your sensory data. This sensory data is merely a stream of electrical impulses that stimulate your computing center.
Bomb #20: In other words, all that I really know about the outside world is relayed to me through my electrical connections.
Doolittle: Exactly!
Bomb #20: Why...that would mean that...I really don't know what the outside universe is really like at all for certain.
Doolittle: That's it! That's it!

Unfortunately Doolittle's plan backfires because although he has convinced the bomb that its detonation orders were faulty, the bomb concludes that all other data is possibly faulty, and ignores new orders to disarm. The bomb decides to detonate anyway - after all, that is what it was built to do.[146]

What Descartes established as the foundation of knowledge, is considered by AI researchers to be grounds for considering a machine intelligent. It is important that the bomb chose to explode. Something that "just runs programs" cannot choose, and so cannot be considered self-conscious. For these reasons, it is misleading to compare human language with computer languages. If human language worked like computer language, the behaviourists would win the day - human beings could be proven to be no more than stimulus response machines. AI researchers have already built "robots" with complex stimulus response patterns - but intelligence involves choice, and stimulus response machines are not exercising real choices. They are choosing from a limited menu of options and the chain of choices will eventually loop as it does in the games and encyclopedias one finds on CD ROMs.

If natural language was anything like a computer language, linguistic relativity would have some credence, and language could be used to embed a representation of the world into people and robots. However, the capacity for self-reflexive thinking, with the freedom that entails, would be missing. As we saw in *Babel-17,* only a programming language can force a view upon a speaker. Language which allows reflective thinking, particularly self-reflective

[146] See Matthew Hurt's Cybercinema Web-Site for an interesting discussion of these issues. http://www.english.uiuc.edu/cybercinema/main.htm

thinking cannot impose any strict world-view, and cannot dictate action. We must conclude that the grammar and vocabulary of language cannot enforce a physics or a metaphysics, and cannot therefore be the basis of a particular model.

This view of language is borne out by Hockett in an essay entitled "Chinese vs English: An Exploration of the Whorfian Thesis", where he proposes a more moderate interpretation of Whorf's theory. Hockett establishes that English and Chinese differ, not in what it is possible to specify in either language, but in what is "relatively easy or hard to specify." He further observes that "from the time when science became observational and experimental...speech habits were revised to fit observed facts, and where everyday language would not serve, special sub- systems (mathematics) were devised."[147]

Hockett continues,

> The impact of inherited linguistic patterns on activities is, in general, least important in the most practical contexts, and the most important in such goings-on as story-telling, religion and philosophising - which consist largely or exclusively of talking anyway. Scientific discourse can be carried on in any language the speakers of which have become participants in the world of science, and other languages can become properly modified with little trouble; some types of literature, on the other hand, are largely impervious to translation.[148]

Hockett suggests that the language of science cuts across the boundaries indicated by Whorf's "linguistic relativity principle," and distinguishes the use of language in a practical context from the use of language in literature.

Hockett's argument entirely undermines the idea that language somehow makes us think in a particular way by establishing that in practical situations the 'limitations' of language are easily overcome. Literature and philosophy, in his view, are more likely to be affected by inherited linguistic patterns because they are not anchored and tested in reality.[149] Scientific language, and mathematics, are largely impervious to such cultural peculiarities because the world of science and mathematics is an international community with agreed methods of testing.

Hockett's argument establishes that equipping our robot with a particular language would not necessarily lead it to a particular view of the world. The vocabulary and grammar of a language would not prevent it from expressing

[147] Charles F. Hockett. "Chinese vs English: An Exploration of the Whorfian Thesis" Ref?
[148] Ibid.
[149] Derrida views this leading astray as a vicarious kind of virtue when pushed to extremes.

certain views of the world. A restricted vocabulary and grammar would only make it more difficult, but not impossible, to say certain things. Hockett also attacks Whorf's argument from another direction.

Language and Culture

Whorf's method assumes that one can make generalisations about a people or a culture based on observations of the vocabulary and structure of their language. Whorf attributes mental habits, thoughts even, to people based on peculiarities of their grammar. He presents us with a circular argument which holds that the language of the Hopi, being the product of minds and of a culture which is foreign to the westernised reader, predisposes native Hopi speakers to a world-view radically dissimilar to our own. He then suggests that it is language which entraps the Hopi in this world-view, and that we as native English speakers are similarly trapped, all unaware, in the world-view of our own language.

Whorf's methodology is seriously flawed, and his basic premise is mistaken. The study of the grammar of a linguistic community will not bring the student any closer to understanding how that native speaker thinks, and it is fatal to regard language as a reflection of mental activity. Whorf believes that such study provides insights into basic mental operations of native speakers. In fact, all that it is likely to indicate are some peculiarities in the history and philology of a particular language. Hockett argues this point in his "Chinese vs English" essay through the example of the Chinese word for railroad train, 'Hwoche', which literally translated means 'fire-cart'. Hockett observes that it is not very useful to imagine that Chinese speakers have mental images of fire-carts when they talk about trains. He points out that the word for electric train is 'dyanli-Hwoche', which translated literally is 'electric-power fire-cart'. He goes on to warn of the danger of drawing any conclusions about Chinese thinking from such philological speculations.

> What is apt to be called the "literal" meaning of a Chinese (or other) form in terms of English is very often the poorest possible basis for any judgement. No doubt the childish errors of nineteenth-century European students of comparative semantics stemmed from just such a basis: for example, the oft repeated assertion that the Algonquians can say 'my

> father', 'thy father, or 'his father', but have no way of saying just 'a father', and hence "lack powers of abstraction."[150]

Hockett offers examples of a series of other grammatical differences and similarities, including a brief look at "handling of space and time", and concludes that to all practical purposes it is impossible to say whether grammatical forms and speech habits precede or are the result of the Chinese "philosophy of life."

> (I)f there is indeed a determinable correlation, then it would impress the writer that the direction of causality in the matter is in all probability from "philosophy of life" to language, rather than vice versa - though, of course the linguistic habit might serve as one of the mechanisms by which the philosophical orientation maintains its existence down through the generations.[151]

An exercise designed to determine the “philosophy of life” of an English speaker might begin with an analysis of the vocabulary and grammar of an educated English speaker. One might assess the percentage of active verbs, or Latinate nouns, one might even look for common metaphorical constructions. Note the spatial metaphors in the following passage.

> He’s *on top* of the situation, in *high* command, and at the *height* of power in having so many people *under* him. His influence started to *decline*, until he *fell* from power and *landed* as *low man* on the totem pole, back at the *bottom* of the heap.[152]

The speaker of this passage, the linguistic relativist concludes, thinks largely in terms of spatial relations. In my view the inference is invalid. One cannot infer the mental habits of an individual from the particulars of vocabulary, grammar or metaphor which the speaker exhibits. These particulars are often part of the

---

[150] "Chinese vs English," p.121.

[151] Ibid. Hockett comments on an unwillingness to identify links between language and philosophy in western languages, and comments,

> The most precisely definable differences between languages are the most trivial from the Whorfian point of view. The more important an ostensible difference is from this point of view, the harder it is to pin down. (“Chinese vs English” p.132.)

Hockett's examples prove that it would be misleading to impute any kind of mental processes, or operations of logic to Chinese speakers on the basis of accidents of the structure of Chinese itself. The grammar and vocabulary of a language do not somehow embody a world-view. We can only get to know someone’s world-view by conversing with them and finding out about them and how they live and how they see the world.

[152] Taken from Richard E. Cytowic, *The Man Who Tasted Shapes.* London: Abacus, 1994, p.208. Cytowic uses this and other examples to argue that metaphors are largely derived from actual bodily experience, and that their coherence comes from this concrete experience and not abstract reasoning. He concludes that metaphorical conceptual thinking determines that we primarily relate to the world emotionally and experientially, not rationally. Cytowic’s analysis is another case of reasoning from language to mental habits.

system or the social milieu, and regardless of education, they cannot be linked to individual thought patterns (whatever that might mean). Furthermore, it is futile to attempt to characterise the social milieu through analysis of the language as a whole. Take *Roget's Thesaurus*[153] as a handy categorisation of the English language. Roget divides English into six major categories:-

Abstract Relations

Space

Matter

Intellect: the exercise of the mind

Volition: the exercise of the will

Emotion, religion and morality

If linguistic relativity were true, and if Barthes and Derrida's complaints had any merit, one might conclude that Roget's analysis would tell us something about English speakers, the limits of English, the tendencies of English to lead us into thinking in a certain way etc.. 3% of *Roget's* concerns Matter, and 22% concerns Abstract Relations. Are we to conclude that English speakers are not materialistic, that they are abstract thinkers, that English leads us to undervalue the concrete? Or in the context of our analysis of the frame problem, conclude that one should not concentrate on feeding AI machines names of objects and descriptions of states of affairs, but develop their perception of abstract relations.[154] It cannot sensibly be said that the grammar and vocabulary of English map onto structures in the world, or on their own tell us anything about the culture of English speaking people.

From a linguists point of view Whorfianism and Chomskianism couldn't be more distinct. Both however argue that language is in some way railroading thought. Whorf argues that our native language determines our world-view, Chomsky argues that deep grammar - which is common to all language speakers - delimits how we can know the world. Whorf seems to leave the door open to different views of the world, whereas Chomsky seems to slam them shut. What is not clear in both these accounts is exactly what they are

[153] *Roget's Thesaurus*. Abridged by Susan M. Lloyd, Harmondsworth: Penguin, 1984.

[154] This may well be a good strategy, but an analysis of Roget's is not a good basis to conclude this. For the record, the percentages for each category are Volition (22%), Abstract Relations(18%), Emotion, Religion and Morality(18%), Intellect(15%), Space(14%), Matter(13%).

leaving the door open to, or shutting the door on. Janlert is more specific, he believes that in order to solve the frame problem the metaphysics of the world needs to be embodied in the form of the representational system used by the AI.

> The metaphysics actually used by human beings in a commonsense world, I will refer to as the operative metaphysics.
> If one believes (as I do) that the operative metaphysics is at least partly reflected in natural language, linguistics offers a wealth of evidence for contrary hypothesis about our fundamental operative concepts of time and change: Rather than latitudinal, absolute entities, one finds longditudinal, relative entities.[155]

Janlert is suggesting a commonsense metaphysics as a replacement for the snapshot metaphysics of most AI modelling systems. Most of the systems which Janlert deals with in his essay (GPS, STRIPS, PLANNER, TMS) rely on tracking changes from one situation in time to another situation in time. Very little consideration is given to time itself. His appeal to linguistics suggests that some kind of commonsense operative metaphysics can be found in natural languages. Furthermore, he believes that this metaphysics can be embodied in the form of the language of representation which the AI has. Thus the metaphysics is a capacity of the system rather than explicit knowledge.

> Ideally the metaphysics is built into the system, so that it becomes embodied in the form, or medium, of the representation. The system simply obeys it, without being aware of it. The metaphysics is then intrinsically represented. [156]

Zenon Pylyshyn in his paper “Rules and Representations: Chomsky and Representational Realism”, explores what is means for hypotheses about the natural world to be “internally represented”, explicitly or implicitly, by a series of rules. He is concerned with what it means for a series of internal rules, equivalent to Chomsky’s Universal Grammar, to map onto the world. He argues that “beliefs must be encoded by systems of symbols which have a constituent structure that mirrors the constituent structure of the situation

[155] “Modelling Change - The Frame Problem” p.34
[156] Ibid. p.37.

being represented."[157] Pylyshyn admits that progress is slow, but believes that the whole future of Cognitive Science hinges upon being able to discover "cognitively impenetrable basic capacities", by which he means properties attributable to the central functional architecture of human cognition which represent internalised knowledge of certain constraints that hold in the physical world. You may recognise this as a restatement of the frame problem. He provides an example of how when interpreting a 2D image as a 3D layout, human beings make a series of assumptions about the natural world which constrains how that picture is interpreted. These interpretative assumptions that are built into our visual system ensure that we don't have to wrestle with lots of different possible interpretations of the image. It is proving difficult to build such an ability into AI machines. Pylyshyn's faith that the answer to this problem lies in some Chomskyan solution has yet to be vindicated.[158]

Pylyshyn would say that the characteristics of vision which have been handed to us by evolution are "cognitively impenetrable", and those that arise from cultural constraints are "cognitively penetrable". It is clear that the kind of language constraints that Barthes and Derrida and Whorf complain of are in the latter category, and those that Chomsky deals with are in the former. Dennett argues that the distinction between these two is irrelevant. If there is a grammar of language which guides our mental activity, why should it matter whether this grammar is genetic or culturally determined? He writes,

> [T]he very vocabulary at our disposal influences not only the way we talk to others, but the way we talk to ourselves. Over and above that *lexical* contribution is the *grammatical* contribution. As Levelt points out (1989, sec.3.6), the obligatory structures of sentences in our languages are like so many guides at our elbows, reminding us to check on this, to attend to that, requiring us to organise facts in certain ways. Some of this structure may indeed be innate, as Chomsky and others have argued, but it really doesn't matter where the dividing line between structures that are genetically

[157] Zenon Pylyshyn, "Rules and Representations: Chomsky and Representational Realism", a paper delivered at the conference "The Chomskyan Turn", Tel Aviv and Jerusalem, April 11-14, 1988.

[158] See "Is Vision Continuous with Cognition? The Case for Cognitive Impenetrability of Visual Perception" by Zenon Pylyshyn, Rutgers Center for Cognitive Science, 1997. In the paper Pylyshyn forwards the view that "certain natural constraints on interpretation, concerned primarily with optical and geometrical properties of the world, have been compiled into the visual system."

> deposited in the brain and those that enter as memes. These structures, real or virtual, lay down some of the tracks on which "thoughts" can then travel. Language infects and inflects our thought at every level. The words in our vocabularies are catalysts that can precipitate fixations of content as one part of the brain tries to communicate with another. The structures of grammar enforce a discipline on our habits of thought, shaping the ways in which we probe our own "data bases", trying like Plato's bird-fancier, to get the right birds to come when we call. The structures of the stories we learn provide guidance at a different level, prompting us to ask questions that are most likely to be relevant to our current circumstances.[159]

It is significant that Dennett puts "thoughts" and "data bases" in inverted commas. Because he is praising language not denigrating it. He is admitting that there may be grammatical and lexical characteristics of language which guide us in our thinking, but significantly, he does not say what they are. The linguistic structures Dennett is referring to hold the key to thinking - but how do the structures relate to what we can think? Chomsky, quite wisely, has declared it beyond our capacity to identify the mechanism whereby we can correlate these two. He has spawned an industry which seeks to identify universal grammar in our genes, but will not say how that grammar can circumscribe how we can know the world. Chomsky is wise to be agnostic on this issue, because universals are tricky things to identify, and it is almost impossible to draw any conclusions once they are identified. Levi-Strauss spent his life examining the structures of myths of

Memes were posited by Richard Dawkins as the carriers of ideas. Dennett refers to them as "good tricks" which is a very good way of thinking about it. Genes carry biological aspects from generation to generation and those that help the organism to survive are more likely to be transmitted - the point being that it is a gene centred system. Genes exist to replicate more genes - organisms such as fish, cats and human beings, are mere vehicles. Dawkins claims that this principle is an evolutionary principle not restricted to carbon-based forms. Memes are the intellectual equivalent, we (and computers and books and other information replicators) are vehicles for ideas - they replicate and use us to do it. As Dennett puts it "A scholar is just a library's way of making another library"

[159] *Consciousness Explained*. pp.300-1.

different cultures in order to identify cultural universals. His quest was inconclusive. One of the problems with identifying universals of any kind is that it is difficult to say which characteristics are accidental and which are integral. Either way it is impossible to say which are significant. Something which we observe to be an integral universal, that is, a universal which is defining for the phenomenon, might also be of no use in making comparisons. One might, for example, observe that all human languages have vowel-sounds, but I am not sure where that gets us. One might also observe, as Hockett does, that all human languages have semanticity.

> Semanticity. Linguistic signals function in correlating and organizing the life of a community because there are associative ties between signal elements and features in the world; in short, some linguistic forms have denotations.[160]

If indeed this is a linguistic universal the quest for a kind of iconic mapping between the structure of language and the structure of the world is horribly misguided. Human beings are not honey-bees doing a bee-dance (see Appendix A), the grammar of our language is not determined by what we want to say. The bee-dance language can only say in what direction and at what distance a certain kind of pollen or food is - that is the nature of its grammar. It cannot be adapted to informing the hive of an approaching enemy, for example. If the bee-dance could shift idioms in this way it would certainly be a candidate for a fully fledged language. Conceivably we could develop a bee-dance type language for our robot where elements of grammar had correlations in the likely operation in the worlds which are to be represented but this would not be a human-type language.

What we can say about human language universals does not tell us much about human thinking. Certainly it is difficult to map the fact that all human languages have first and second person singular pronouns (a speaker and addressee) onto the world.[161] I have known children that do not use “I” or “you” until school age. For example, the following sentence addressed to me by 5 year old Jessica. “Jessica wants to show Ron a picture.” The more usual would be “I want to show you a picture.”

---

[160] C.F. Hockett. “Universals in Language” in *The View From Language*.

[161] I can accept that plural pronouns have a function. If I am addressing a football team I do not want to have to list all their names each time I refer to them. I simply say “you” (plural).

In the Whorfian and the Chomskyan case it is not clear what an intrinsic metaphysics really consists of. In both cases it is assumed that a series of grammatical rules embody a “world-view” which maps onto the world - this, I believe, is a crucial error. Both Chomsky and Pylyshyn sweep aside Wittgenstein’s observation that it is never possible to decide which rules are being followed by a person - or if any rules are being followed at all. To some extent they are misled by the fact that in a computer system it is very easy to establish which rules are being followed. Wittgenstein’s point is that there are many different ways of characterising how one follows a rule. That is, the way one follows a rule when multiplying numbers, might be very different from the way one follows the rules of chess. Crucially, all rules are open to interpretation. Chomsky and Pylyshyn are looking for internalised rules in the human cognitive system that invariably determine certain behaviour and structure language. In AI this principle can only be realised in idealised systems, or in remarkably simplified block-world systems. The languages and principles that have been developed in these systems rarely translate to other systems (they are environment specific) and never to worlds where things can be different shapes, sizes and colours.

For our purposes, if we are going to equip our robot with colour vision and a schemata for identifying and describing the colours of things, where do we begin? - with a dictionary of colour words? *Roget’s Thesaurus* devotes a puny 3 pages to colour - less than a third of one percent of the volume! If vocabulary was a yardstick of cultural significance, colour would clearly be a minor issue for English speaking peoples. In fact, vocabulary is not a yardstick, and colour is not a minor concern. Colour-vision is very important for human beings. Primates are the only mammals with colour-vision. Other mammals, such as dogs and cats, have monochrome vision, but their other senses are vastly augmented. We have discussed equipping our snack-robot with smell and taste, does it also need colour vision? If we enable our robot to see in colour, we will eventually have to describe coloured objects to it, and elicit colour descriptions. Is the language in which we do this going to structure how it can know the world? Is it going to structure how we can understand its world? The following analysis of colour-blindness demonstrates that the answer to both these questions is “no”.

Seeing the World in Colour

Because I am colour-blind, the colour words that most people use do not suit the way that I see colour - there is definitely a problem here. There are a number of colour words which describe colours I cannot see, and in a great many cases I use the words wrongly. This led my teachers at school to conclude that I was either stupid or needed lessons in colour words. It wasn't until I was 13 years old that I was diagnosed as colour blind. I am someone who has *prima facia* case for distrusting language - colour language does not fit the way I see the world. Or to put it in more philosophical terms, my subjective experience of colour doesn't match the colour-word system. Barthes complains that he must construct himself as a subject, and as either masculine or feminine, in order to speak in French. I complain that everyday I mis-describe a colour because colour language does not work for me. Do either of us have a case? Are the cases comparable? Is language the villain of the piece?

Barthes complains that there is an ideological bias built into French terminology. Derrida complains of a similar problem and sets himself the task of revealing the contradictions in oppositions such as subject/object, male/female. Clearly both believe that language projects a structure on the world which they are more or less forced to use when they speak. It would be fortunate for AI researchers if this could be proven to be the case. They don't complain about the deterministic quality of the grammar of natural languages, AI researchers would like to celebrate this elusive quality. AI researchers are hopeful that natural language maps onto the world such that its structure reflects the world of objects in dynamic relationships. They want language to be able to determine how a robot thinks and sees and knows the world.

We know that different languages have slightly different ways of breaking up the spectrum. Following Whorf's line of reasoning, a language which didn't have a word for a certain colour, or was abundant in words for the same colour, e.g. white, would reflect particular blindnesses or sensitivities in that culture or race. We will see that this is not the case.

The English system of colour words presupposes oppositions and similarities which I cannot discern. When I describe the car as blue, I find that it is purple. If "blue" and "purple" were the same word, "bluple", that would fit the way I see things. Unfortunately, only colour blind people can really understand what I mean

here. I seem to be claiming that there could be a kind of private language between colour blind people - I am! I could devise a series of words which would enable red/green colour-blind people to describe the world to each other without contradiction. It would be an easy matter to develop a substitute vocabulary such as with the above case of replacing "blue" and "purple" with "bluple". But suppose I wanted to develop a grammar that reflected how colour-blind people see the relationships between colours? This structural and grammatical task is the one that AI researchers need to tackle if they believe that language can make relationships implicit to the user. In the case of my own vision, I wouldn't know where to begin. Pylyshyn admits that this is the kind of instance where cognitive science cannot come up with correspondence rules.

According to colour theory, green is the opposite of red. However, for me, they are often indistinguishable from one another. Pinks look grey. Maroon looks black. Clearly there are different colour relations here, but I am unable to codify them into a structure. In looking for a structure, I inevitably ended up studying colour theory and I now know more about it than most normally sighted people. I can mix a sky blue, or a sea green with no difficulty at all - as long as the paint tubes are labelled! I am able to guess the colour of something because I understand how my vision relates to the spectrum of colours, and I understand how colour-words map onto this model. For example, I know that I mistake browns for greens because I am insensitive to red, and brown is a mixture of red and green.

My limited success as a colourist and my ability to circumnavigate my problem indicates that my eyes are responding to phenomena in the world, but are unable to respond to in the same way as normal eyes. I have, by adopting the colour theory model, used the universality of those phenomena coupled with a "scientific" approach, to deal with my problem. This is a powerful argument for equipping our robot with a scientific account of colour. Science regards colours as electromagnetic radiation of various wavelengths. If normally sighted people could identify colours using wavelengths wouldn't life be simple? When you wanted to repaint the woodwork you would merely have to tell the paint shop which wavelength and saturation and hue you required - no more colour

swatches.[162] The scientific account of colour indicates that for me, certain wavelengths look the same - it is as if I couldn't see the difference between a 6ft bench and a 4ft bench. In fact it is probably better to imagine my vision being insensitive at one end of the spectrum - in much the way that human hearing is truncated at pitches that dogs can hear. If we equip our robot with infra-red and super-hearing, a scientific model is going to come in very handy.

If I had only the system of colour words without the supplementary scientific account of colour, my colour vision would remain a puzzle to me, and I would be far less successful at guessing colours and mixing paints. In this sense, it is possible to say that the colour word system alone is inadequate. This is what Hockett means when he argues "where everyday language would not serve, special sub- systems were devised." Does this mean that "everyday language" is the villain of the piece? Should we adopt Bertrand Russell's approach and try and devise a perfect language? I am sure many AI researchers are attempting that very task . They will fail.

A note on Bertrand Russell, symbolic logic and logical atomism.
Bertrand Russell came to the conclusion that many of the problems of philosophy were both engendered and hindered by the fact that they were debated in "ordinary language". In order to get around this problem, Russell imagined a perfectly logical language without the redundancy and lack of precision which characterises everyday language. The logical atomists believed that such a language could, theoretically, provide a complete description of the world.

If I run a red traffic-light, I do not say to the constable, "I ran the light because your colour words don't work for me." I ran the light because I cannot distinguish the colours that have been chosen to signal stop and wait. If red signaled stop, and blue signaled wait, I would have no problem. Colour-words are not the problem here, but the choice of colours used to signify stop/wait/go is a problem. If the system of traffic lights is regarded as a language (which, following Wittgenstein's examples in *Philosophical Investigations* it no doubt could be), then language is the problem. The system makes a different series of distinctions from those that I make. The system assumes that everyone makes those distinctions, and those distinctions are hard-wired into the system. From this perspective it seems that Derrida and Co.

[162] In fact, that would only be true if lighting conditions remained the same everywhere.

have a case - language disadvantages those who do not perceive the world in a normative way because it reflects and enforces the majority view. They argue that the varied subjective experiences of different people are being given the appearance of homogeneity by the language in which they must express these experiences. My subjective experience of colour is different from that of most other people yet I must still express my colour experience using a system of words which does not fit my experience.

These "subjectivists" argue that everyone's subjective experience of colour might be different, and that the system of colour-words guides them toward agreement only about the public phenomenon. This puts the colour-blind in a larger class of people whose subjective experience of colour might be varied and exotic. The subjectivist argument is that the internal experience of colour vision may vary radically from person to person.[163] I am not sure that this makes sense. Obviously my subjective experience of colour is very different from the majority of people - my behaviour shows that I cannot distinguish colours that other people can - red and green look the same to me. It was not just the way I used language which originally highlighted this problem - it was the way I accelerated up to traffic-lights! In the absence of traffic-lights or language, not being able to distinguish ripe and unripe fruit might have brought the problem to the fore. When a person makes a colour mistake, it is reasonable to conclude that that person has a different subjective experience of colour from normally sighted people. But it is that person's public behaviour that is important. It is not useful to speculate on the nature of the person's internal colour state. Speculation on a colour-blind person's subjective experience of colour could lead one to make the kind of assumptions Whorf made about the Hopi. That is, it might lead one to conclude that colour-blind people could discern no colour at all, or that they weren't fit to look after children, drive cars, fly aircraft etc. When people ask me how the world looks to me I'm often at a loss because it all looks perfectly normal to me. It is only when someone sees something I can't - a red ball against a brown field - that there is a subjective experience to relate. What I am saying is that if we need public evidence in a case where we know that a person (me) has a different subjective experience, then we will need public evidence in the case

[163] This line of thinking has given rise to the famous "inverted spectrum" problem.

where we are merely speculating that people's subjective experience of colour is different. If you could suddenly see the world through my eyes, you would notice a difference from what you normally see. You would be surprised and fascinated for a while, but these are your subjective experiences, not mine. You would, like me, fail to see the red ball in the brown field. Congratulations, you just failed to see something that I failed to see - does that mean we are having the same subjective experience? The public evidence that our colour-blind robot has a different subjective experience is that it runs a red light - who is to blame, the robot or the traffic-light system? If the language system of traffic-lights is the villain, then there seems to be a case for those who claim that language can enforce normative views. Some, of course, would argue that that is exactly what traffic lights should do. Some communication systems, such as traffic lights, air-traffic control systems, bar-coding, are designed to be as unambiguous as possible. If our robot's responses were coded using a computer language it would be unable to act in any way other than that in which it was programmed. Its "choices" would be limited. As we have seen, if our robot understood phenomenology, it could decide, like bomb #20, on its faulty evidence, whether to stop or go, or even do something else entirely.

These reflections on colour-language indicate that natural language has characteristics that disqualify it from the capacity to impose a model, to carry a model within its structure, or to in any way facilitate the inculcation of a model. The power of language to enable complex thinking is not counterbalanced by the kind of limits and restrictions which many philosophers imagine. The limits and restrictions are imposed by the culture of which language is a part. If one has difficulty imagining another way of living, a different kind of mathematics, or even alternative intelligences - it is not the fault of language but of the culture you were raised in. Language didn't determine that culture, but it helped enable it. There is a crucial difference between a determinate process which establishes a definitive state of affairs, and a series of enabling factors which make a number of solutions possible.

Installing a colour language in our colour-blind robot is not going to enable it to distinguish colours, although it may make it aware that it can't distinguish colours. By the same virtue, installing such a language in a colour-sighted robot will not enable it to see any relationships it cannot already discern. It may

facilitate communication about colours, it may even make it aware of colours outside the visible spectrum, but there is nothing in the structure or vocabulary of colour which has enabled this. Those who attribute perceptual abilities and restrictions to syntactic structures in natural language forget two important characteristics of natural language - redundancy, and flexibility of categories.[164]

Redundancy in Language

Even though I am colour-blind I rarely run red lights. Why? Because the system has a fail-safe - the position of the light. Although I cannot distinguish red from amber, I know that when the top light is lit I should stop. Given the number of colour-blind people in the world, it would seem that if people stopped at lights only on the basis of colour, life would be a lot more hazardous at traffic lights. The system works because of what normally sighted people might call the "redundancy" in the system. I have no doubt at all that some economic rationalist somewhere is at this moment devising a traffic light system with only one light. It will fail because it will not work for 10% of the male population.

It is the redundancy in natural language, and especially English, which makes it so adaptable. The traffic light system is a successful language because it shares this quality of redundancy with natural language. When the economic rationalist develops a new traffic-light system, it will no longer have redundancy, there will be a lot of crashes, and the system will be to blame. Language systems that are devised for specific purposes are likely to serve specific users, but even something as specific as a traffic light system is capable of serving a wide variety of users.

Redundancy is erased from systems where users are considered to have universal characteristics. The ideal computer program is considered to be the one most economical on code. In fact, the most successful systems have the

[164] Chomsky's argument is that deep syntactic structures enable and delimit abilities and degrees of freedom. His argument hinges on a comparison with coding. Any given code, a bar-code for example, whilst enabling staggering feats of stocktaking, is limited by its length and format in what it can communicate about a given product. The first three digits might indicate country of origin, the next three weight, etc.. Chomsky argues that human languages are structured by syntactic coding which whilst enabling much of human culture, also delimits which ways that culture can develop. (In fact, bar-codes have no meaning at all - they are simply allocated to products sequentially. see *New Scientist*, 25th January 1997, Letters page.)

flexibility to accommodate unforeseen users and data, and long-winded code can often be the most accommodating. In short, if it works it is good. Language evolves, and efforts to rationalise it, such as that made by Bertrand Russell, are usually disastrous because languages developed for specific purposes outmode quickly. Russell's mistake was to assume that there was a world that was essentially the same for all logical beings. For example, who can reasonably deny that twice 2 is 4? Well, the English language does not assume that this apparently universal truth is the case. It is quite simple to question this "truth" using English (Twice two is five.). Mathematical logic is not hard-wired into natural languages.

If language were structured by the economic rationalist developing the single signal traffic light, Derrida and Co. would have a case. Language, however, isn't structured by programmers, or linguists, or teachers, or anyone. By its very nature, language cannot lead us to see particular relationships in the world.

The Flexibility of Language Categories

Some languages lack the range of colour words we have in English. Does this mean that English is a better language, or that these other languages are impoverished? If language lays down the tracks on which thoughts run, is it possible that some languages actually prevent speaker from seeing certain colours?

There is a popular myth that Eskimos have 40 words for snow, a myth which makes English speakers feel impoverished, because in English there are probably only four (e.g. sleet, slush, blizzard, snow). The categories of English are clearly short-changing Arctic explorers. The argument continues that if our language is stunted with regard to snow, there must be other categories where it is deficient. Barthes mentions subject and object, masculine and feminine. Can we conclude that our language has already divided up our world? Are we back to the problem of language enforcing a model?

Many languages discern subject and object, and masculine and feminine, but how many do so in a way that is compelling? The romance languages are full of arbitrary gender assignments which, in fact, undermine gender distinction rather than enforce it. For example, the Italian word for egg, "uovo", is masculine. The

English language has a very relaxed attitude to subject/object distinctions. Language has its categories, but they seem to have a logic entirely their own.

Science describes different colours as different electromagnetic wavelengths, and thus avoids the peculiarities of natural languages. Botany has developed its own system of classifications of plants because of the inconsistent way natural languages, such as Chinese and English, divide up the class of fruits and berries, for example. These divergences are cited as proof that language is categorising the world for us in an insidious way. In fact, when you think about it, every time you walk into a greengrocers, or a fresh food section at the supermarket, you confront an alternative way of categorising the vegetable world. Tomatoes, although botanically classified as fruits, are in the vegetable section. There is a section for nuts and dried fruits, which are not botanically related. The fact that science and greengrocers divide up the world in different ways does not result in an epistemological dilemma - both ways of categorising the world are legitimate.

In this case why should we worry that the Eskimos have such an advanced snow-word system? In fact Eskimos cannot distinguish more types of snow than English people (although they see a lot more of it). The ski report on the nightly weather forecast distinguishes a range of different snow types - some of which even Eskimos might not have words for (examples: powdery, hard and compacted etc.). Note how these descriptions are not always single words. Why should it be significant that one language has a single word for something that in another language requires two words or even a whole sentence? Remember that Hockett's analysis of linguistic relativity suggested that it was unwise to impute something about a culture or people from the grammar of its language. This is borne out by Eskimo language. It transpires that Eskimo words are composite words based on stems. There is no word for snow in general, but four stems which may be paraphrased in English thus:-

snow in the air

drifting snow

snow lying on the ground

soft, watery snow

The myth that Eskimos have 40 words for snow arises because in their language there is one word for "drifting snow in the doorway" (a composite of the stem and

"in the doorway"), for example, and other composite words for "snow on the roof", "snow that you bring into the house on your shoes" etc. In fact, the Eskimos cannot distinguish more types of snow than other people, and their language isn't specialised to do so. It transpires that the categories they have for snow are much like those that anyone might recognise.

We can use language to divide the world up in an infinite number of ways, and even if we can't find single words to express things, composite words and sentences are quite adequate. Those who imagine that we divide up the world primarily with language forget that we directly experience reality when we cut our finger or something hits us on the head. In these cases the pain is not being modulated through a linguistic medium.[165]

4. The Role of Language in the AI Debate

The above analysis of language and colour has established some things which language does not do, and hopefully dampened any suspicions that language imposes some kind of ceiling on our freedom to know things about the world. How does this help us with the frame problem? In essence, I have argued that although language is a powerful tool it doesn't come with instructions on how to use it.

Chomsky argues that the structural parameters of universal grammar are like switches which if set on or off can effect various grammatical peculiarities in individual languages. This is clearly an analogy to genetics, and highlights the meaning independent nature of the structures which Chomsky argues determine our system of knowledge and language.[166]He argues that the switches might turn on one ability or characteristic at the expense of another, and that it is more likely to be the physical structure of the brain and its attendant mechanisms which determine such on/off pairings rather than a semantic, meaning related, or utilitarian considerations. In this way Chomsky side-steps the issue of how language evolved by arguing that the development of the language organ was an accident of evolution. Universal grammar is not structured like any particular human language, and does not reflect structures in the world because the

[165] Dennett doesn't fall for this dualism. His multiple drafts model allows for a whole "pandemonium" of voices and accounts to shape our perception of the world.

language faculty didn't evolve in response to evolutionary and environmental pressure for language. Chomsky suggested that the development of insect wings for example, was equally happenstance.

> Insects have the problem of heat exchange, and rudimentary wings can serve this function. When they reach a certain size, they become less useful for this purpose but begin to be useful for flight, at which point they evolve into wings. Possibly human mental capacities have in some case evolved in a similar way.[167]

The structure of an insects wings might not be reflected at a genetic level, but their structure and shape will be largely determined by environmental factors. The structure of an aircraft at a molecular level is not reflected at a macrocosmic level, but aspects of an aircraft's structure are determined by the molecular structure of the material from which it is made. Its shape is more thoroughly determined by the structure of the air it must fly through - its environment. The same is true of language. Syntactic structures do not determine any semantic features - as with the aircraft, it is the environment in which language must operate which determines its shape. If we take universal grammar out of the equation - even if it existed it couldn't tell us much about the aspects of language and thought that we are interested in - we can see that Chomsky and the linguistic relativists have simply got the direction of causation around the wrong way. As Dennett says, "internal states get their meanings from their functional roles"[168] and not *vice versa*.

My reflections on this issue do not rule out the possibility that in a block-world a code might be developed which will enable a robot to navigate its block environment and achieve some success. This solution to the problem is a Deep Blue solution, and has no connection with how human beings actually solve the frame problem, because a code language cannot be scaled up to become a natural language. Victor Zue, head of the Spoken Language Systems Group at MIT, admits that the problem of communicating verbally with computers cannot be solved by scaling up the computing power of speech recognition systems. In an interview with Scientific American he says,

> In fact there is a danger of throwing computing at the problem and thinking that somehow you will be able to solve it. People think, "If only

---

[166] See *Language and Problems of Knowledge,* pp.61-63.
[167] Ibid., p.167.
[168] *Darwin's Dangerous Idea* p.411.

> I could collect more statistics about how often one particular word follows another." My personal opinion is that our ability to compute is really not the barrier. What to compute--and more fundamentally our knowledge gap about how we communicate--those are the fundamental questions.[169]

Zue believes that real-time dictation with a large vocabulary is only a couple of years away, courtesy of the vast computing power which will soon be available. But this will not solve the hard problem of natural language understanding in machines. Zue does not believe we have enough good ideas about language understanding and higher-level semantics to harness that computing power to solve the problem.

> The problem is that we don't understand human cognition. Something that in humans requires only an almost instantaneous response--to understand the nuances of somebody else's speech--is something that we just don't have a handle on how to do in machines.
> I don't believe for a minute that if you have machines with bigger storage and faster computing that you will be able to solve this problem faster. In fact, researchers in this field have been able to ride the technology curve very well, and we are not hampered by a lack of computing power. We're hampered by our inability to come up with ideas that work. To fundamentally understand human cognition remains a holy grail.[170]

The general goal of AI is not necessarily to understand human cognition, but Zue believes that it may be the means to an end. Dennett believes that the key to human cognition lies in an understanding of how evolution has shaped us.

Human beings, because they are animals, primarily divide the world into pain/pleasure, hungry/full, day/night, friend/enemy and other such universal categories. Language does not supply these categories. In the 20th century, those who accept Darwinism, also accept that much of our basic behaviour is a throwback to our evolutionary ancestors. AI robots have no such ancestors, they don’t experience pain or hunger, and they have no enemies (except maybe those who disagree with AI, such as John Searle). In a way, the basis of our world-view is wired into us, and may form the basis of many of our most fundamental motivations. Above this basic level, the human brain has a hierarchy of other levels, some specialized to certain functions, some not. The

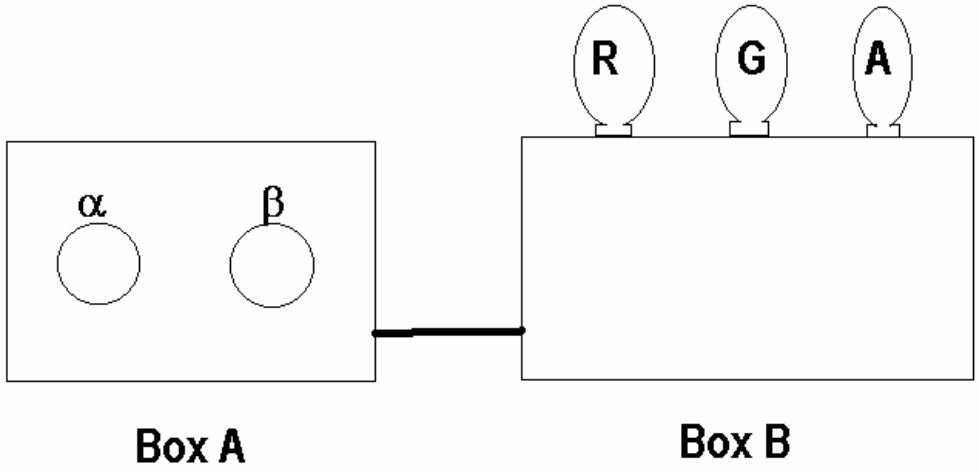
[169] Interview with Victor Zue conducted by W.Wyatt Gibbs, October 1997.
[170] Ibid.

adaptability of the human brain allows us to override our hunger, for example, and carry on working. In Dennett's view, at levels above the basic hard-wired level, we "virtually rewire" our brain, and we do it with language. Evolution and our early upbringing has hard-wired some responses into us. Evolution has provided specific solutions to specific problems. For some animals, when the environment cooled, fur evolved. Where prey comes out at night, the predator develops night-vision, or sonar. These adaptations typically take tens of thousands of years, and are survival mechanisms for specific animals in specific environments. Human beings need to survive in a wide range of changing environments, and Dennett argues that it is the virtual-rewiring enabled by language that enables us to do it. Instead of waiting tens of thousands of years for evolution to come up with a solution, language enables us to try our a number of solutions in our heads, before adopting the appropriate one.

This account of how the human brain works leaves the door open for an evolutionary solution to the frame problem for human beings. It could be that the frame information that enables us to navigate the world is hard-wired at a level that is pre-linguistic. It could be that we learn such things as we grow up. Either way we know that the frame problem is largely resolved in human beings *before* language takes hold. I would even hazard that the frame problem must be resolved before language can be acquired. It will be impossible to teach a robot natural language without first resolving the frame problem because agreement about the frame is the crucial step towards language and communication.

In *Darwin's Dangerous Idea* Dennett revisits the frame problem in a rather bizarre manner, and inadvertently demonstrates how agreement about the frame can form the basis of communication. He argues that consciousness, in evolutionary terms, must have developed like eyes and wings. In his view, human beings are a bit like scaled up frogs - intelligence, consciousness, language are features of complex adaptation. His Two Black Boxes thought experiment is designed to demonstrate how absurd it is to require that the internal workings of an AI machine manifest qualities which we associate with intelligence, such as intentionally and meaning. His main aim is to disable objections that the mind cannot be the product of firing neurons or chemical interactions.

He imagines two black boxes connected by a copper cable. The first box has two buttons, α and β. The second box has three lights, red, green and amber. When button α is pushed, the red light flashes. When β is pushed, the green light flashes. The amber light never flashes when a button is pushed. To cut a long story short, the scientists in the fable are baffled at the complexity of the signals that turn the lights on and off, so they open up the boxes to discover how they work.



The two black boxes in Dennett's thought experiment have sophisticated Lenat-like AIs in them, each with an encyclopaedic common sense view of the world. They are full of obvious facts such as, "Water is wet" and "Down is the opposite of up." They are, in short, full of frame information, and they are in full agreement about how the world is, despite having very different internal architectures (perhaps one is a Mac and the other an IBM). When button α is pushed A sends a message, such as "Butter is food," to box B, box B agrees that this is a true common sense statement, and the red light flashes. When button β is pushed A sends the message such as "Turkeys explode when removed from refrigerators," box B agrees that it is a falsehood, and the green light flashes.

Why has Dennett revisited the frame problem in this bizarre way? Wittgenstein argues that language is not based on agreement in *terms* but on

agreement in *judgements.*[171] Dennett's two Black Box AI's agree in their judgements about the world because they have been fed the same frame information. In this scenario, the frame information is a pre-requisite for communication between the boxes. Human beings must largely agree about how the world is before we can communicate. A frog is given its frame by evolution. These Black Box AI's were given their frame by a couple of human "hackers". The reason a frog can survive is because its evolution-given frame information is eminently suitable for the environment it must survive in. The Black Box AIs' frame information is true only for an Earth-based environment. If you put them in the space shuttle they will continue to agree about a whole lot of things that are no longer true. If their survival depended on it they would die. A frog in a space shuttle would probably die too. Frame information is a pre-requisite for language but it is not fixed, and the flexibility of language and its categories ensures that frame and language evolve together.

Dennett's Black Box example is also a dramatisation of Wittgenstein's dictum "Meaning is use". Dennett argues that meaning is not an intrinsic property but derives from function. The bottom line is that the kind of deep meaning which human beings find in the world is merely a scaled up version of the kind of meaning a frog finds in its world. Different species see the world in different ways because what is significant is different from species to species. Something that means one thing to one species, means something else to another. This revisits Wittgenstein's comment "If a lion could speak, you would not understand it." Because differently embodied creatures will be led to different views of the world, communication between species, and by extension communication between man and machine, will be limited and will fall short of the kind of communication which human beings experience. If other species and AI machines developed something like natural language, which enabled them to become autonomous, even conscious entities, they would think in a way so unlike human beings that we would hesitate to call them intelligent.

If the frame problem is a model problem, it is clear that the kind of model of the world that a robot can have is very different from the kind of model a

[171] *Philosophical Investigations. passim.*

human being might have. The way the robot sees is very different from the way a human being sees. If a human being always sees things as something, as Wittgenstein suggests, it is tempting to think that human beings must always be trying to match what they see to a kind of internal database of images and a model. This likens us to sophisticated information retrieval machines. Much of the work being done building recognition systems for robots uses this principle. The robot looks at an object and compares it with its database until it finds a fit. For example, if a robot is waiting at a bus stop, it hallucinates that everything that comes over the horizon is a bus. It summons up an image of a bus and compares it with what it sees on the horizon. It is tempting to think that human beings recognise objects in this way, but recent research shows that this is not the case. When you think about it, when we search our brain for something, like a name, or a face, we don't start with the thing we are looking for. Often we have no idea what the name is, yet we know it when we find it. Computer databases do not work like that. They start with a name and find a record which matches it to a phone number or address. The recognition system which these particular AI researchers are developing throws very little light on how human beings recognise objects. Victor Zue has no such illusions about his speech recognition system - it recognises speech, but it does it in a computer way, not a human way. It can recognise a grammatical sentence and respond in an ELIZA-like way - this doesn't scale up to a system which understands natural language. The chief method used in getting an AI to recognise change is to ensure that it has exhaustive knowledge-base about its environment which it can revise when it discovers that it conflicts with the real world - in most cases a block-world. Brooks has found evidence that this is not the way human beings recognise change.

> There is evidence that in normal tasks humans tend to minimize their internal representation of the world. Ballard, Hayhoe & Pelz (1995) have shown that in performing a complex task, like building a copy of a display of blocks, humans do not build an internal model of the whole visible scene. [172]

[172] Rodney Brooks, Cynthia Breazeal (Ferrell), Robert Irie, Charles C. Kemp, Matthew Marjanovic, Brian Scassellati, Matthew Williamson. "Alternate Essences of Intelligence" (Submitted to AAAI-98), January 1998.

Once again we find that in order to solve the frame problem AI has resorted to a method which is computer-bound and throws no light on human cognition.

The frame problem for a robot builder is simply making the robot recognise change. The frame problem where it concerns human cognition is how we know even the most basic facts about the world such as: situations change as a result of actions; and when a thing is in one place it can't be in another. I dismissed the idea that language structures the way we can know things about the world. I dismissed the idea that language can circumscribe what can be said. The frame problem is not a language problem, AI researchers who believe this have cast the problem the wrong way round - the language problem is a frame problem.

One forgets that a great deal of stage-setting in the language is presupposed if the mere act of naming is to make sense. Ludwig Wittgenstein[173]*.*

Language-games

What Wittgenstein calls the "language-game" is the context whereby meaning arises through language. Every situation involving language can be considered a language-game which more or less creates its own rules. A conversation with the postman; a lecture on physics; singing in the shower; each of these is a language game whose rules are roughly determined by similar language games. The same phrase used in each of these language games, for example "It's a lovely day today" will have radically different meanings. There are an infinite number of language games although a number of philosophers and linguists have attempted to break them down into types.

In Wittgenstein's view, most of the puzzles in philosophy arise from a confusion about which language-game one is in. Philosophical paradoxes are, in his view, chiefly brought about by an attempt to use the language of one language-game in the context of another.

Frames and Language-games

The frame problem needs to be solved before the language problem because establishing the frame is an important step towards language. As Wittgenstein suggests, the stage must be set before the players in the language-game can be understood. The words of language only have meaning by virtue of the role they are ascribed in the language-game. Teaching a computer the names of objects and providing it with a few grammatical rules, in order that it can make sentences, is an empty game unless pre-requisite aspects of the particular frame are established.

[173] *Philosophical Investigations.* 257.

The “brittleness” of computer knowledge is a function of the knowledge being specific to a frame, or area of expertise. We saw this with the example of the medical expert system which diagnosed a rusty car as having measles. But it is not just knowledge, in the sense of brute information, that makes each frame “brittle”, it is the fact that what counts as imitation, accuracy, rule-following, and a host of other basic frame functions, vary from area to area, from frame to frame, from language-game to language-game. Lenat is rapidly discovering with CYC that each frame, or microtheory, is another language-game which has its own set of rules. What he will eventually discover is that human knowledge is less brittle than machine knowledge because human beings grow up learning a whole range of inter-related activities, each with its own rules, goals and practices, and, according to Wittgenstein, the way language fits each activity is unique - it is a language-game.

Because language connects so uniquely with human activities e.g. giving orders, coaxing, lying, asking questions, and because each of these basic activities varies according to social setting and specific context, it is impossible to apply invariable rules which will cover more than one language-game. This has clear applications for AI research. A block-world will be amenable to a block-world language solution, but the solution might not necessarily transfer to other block worlds or to the world at large.

The language which suits a particular frame cannot be learnt except within that frame, through participation and association with activities. As we saw with Data, imitating movements and sounds does not constitute learning. If there is a rule to be followed e.g. a multiplication table, dance steps, winning poker hands, it is probable that what constitutes following a rule in each case will be determined by different criteria.

In this sense, following a rule is not a simple activity which can be defined by saying that the person did what the rules specify.[174] In each case the point at which one can say that someone followed a rule will be determined by different criteria.

[174] For a more thorough discussion of Wittgenstein’s notion of rule following see Saul Kripke, *Wittgenstein on Rules and Private Language*. Oxford: Basil Blackwell, 1982. See also http://csmaclab-www.uchicago.edu/philosophyProject/chomsky/Kripkenotes.html

Learning a language is not simply rule-following as Chomsky suggests. It is a complex of activities in which a number of notions of rule-following hold sway. This view of language is supported by the work of Patricia Greenfield of UCLA, who initiated a study of how her own children acquired language and compared it with how apes developed.

> I knew all about the Chomskyan approach to child language development, which is an approach in which grammar is very central, and the child is considered sort of like a little grammar machine, or becoming a big grammar machine. And when my daughter Lauren started to speak, what absolutely hit me was ... what she was doing was nothing like what they were describing. And in fact, what they were describing were children combining words with words, using rules, but what she was doing when she first started to talk was combining words with things, with people, with gesture, all sorts of non-verbal elements.[175]

Greenfield exposes the weakness of the Chomskyan model when she observes that one can generate neither grammar nor meaning without a social context. It is not enough to combine words with grammar, or even words with things, one must learn the uses of words (and grammar) as one combines them with human activities. In Greenfield's view, being embodied in the world has a grammar of its own, and language is learnt simultaneously as this body-grammar takes shape for the learner.

Earlier we established that differently embodied creatures would arrive at different views of the world. They will operate in the world differently and their communication needs and means will be different. Various aspects of how a creature reacts with its environment will trigger the appropriateness of certain language-games and not others. It is not surprising, therefore, that embodiment is crucial to language learning.

This conclusion has a downside for how one assesses the intelligence of creatures other than ourselves - and that includes machines. The problem with the artificial intelligence debate is, as Robert Wright says in his *Time Magazine* article, "Can Machines Think?", that every time a criteria of intelligence - the Turing test, beating a chess master - is surpassed by a computer, the goalposts are moved. A new test is suggested. What we should learn from this is that there

[175] From transcript of a Nova program at Journal.Graphics@bobj.cc.uic.edu Newsgroups: jrnl.pbs.nova Subject: [1/5] Can Chimps Talk? Date: Sat, 19 Feb 1994

are no definitive criteria for intelligence. Intelligence tests just measure how good we are at passing intelligence tests.

Moving the Goalposts

This goalpost moving is not confined to AI research. The arena in which it shows itself most clearly is in the development of computer chess-playing programs. A decade ago hardly anyone believed that a computer would ever be able to beat a grand-master. Today, with Deep Blue, it is a reality. Even if Deep Blue consistently defeats grand-masters, there will be those who argue that playing out duplicates of strategies and games stored in a database is not playing chess. I am sure that Dennett would agree that this is the case. He points out that the goals of chess programs and AI programs are very different. Yet he warns of goalpost moving within AI itself. In fact, what we find in AI are a number of research projects, each developing artificial intelligences, but with widely divergent views of what constitutes success. Are they modelling human cognitive processes, or extending them? The list of reasons for making robots in humanoid form, or computers that speak with a human voice, or model human cognitive processes, makes interesting reading. Dennett believes that despite the fact that most AIs are made up of "cognitive wheels" - unbiological design elements mimicking human cognitive processes - there is still a lot they can teach us about human cognition.

> Someone who failed to appreciate that a model composed microscopically of cognitive wheels could still achieve a fruitful isomorphism with biological or psychological processes at a higher level of aggregation would suppose that there were good a priori reasons for generalised scepticism about AI.[176]

In short, if you don't believe that a suitably sophisticated digital machine (for example) could usefully model human thinking at a high level, then you don't believe that AI can achieve anything. It ought to be noted here that some AI researchers see AI as an engineering problem to extend, not emulate, human cognitive powers. In Dennett's view, AI is useful because it asks hard "How do we do it?" questions about human cognition, and asks,

[176]"Cognitive Wheels," p.148.

> Even if on some glorious future day a robot with debugged circumscription methods maneuvered well in a non-toy environment, would there be much likelihood that its constituent processes, *described at levels below the phenomeno-logical*, would bear informative relations to the unknown lower-level backstage processes in human beings?[177]

His answer is that AI builders should beware of solving problems using "cognitive wheels" - unbiological solutions. Clearly an AI model cannot exhibit the features of a human mind at all levels - after all we are not talking about building biological brains.

> No one supposes that the model maps onto the processes of psychology and biology *all the way down*. The claim is that for some high level or levels of description below the phenomenological level (which merely sets the problem) there is a mapping of model features onto what is being modeled: the cognitive processes in living creatures, human or otherwise.[178]

The Turing Test
In his 1950 paper "Computing Machinery and Intelligence" the AI pioneer A.M.Turing proposed the following test (which he called the "Imitation Game") for whether a machine can be considered a thinking machine.
A computer and a person are placed in a room, and a third person, the interrogator, in a separate room. The interrogator communicates with the two entities in the room via typescript and must guess which is the computer.
Turing proposes that if a computer could pass this test then it could be considered a thinking machine.

There is a lot of room for argument and manoeuvring (and goalpost moving) when we begin to talk about the levels at which an AI should model the cognitive processes of living creatures. At one time, if a computer merely fooled human beings that a mind was at work - by engaging them in conversation, or beating them at chess - that was enough to herald a new age of sentient robots. It would make matters a lot less acrimonious if each project had drawn a line which an animal or machine must cross in order to qualify as intelligent. Unfortunately, that has not been the case. In Dennett's terms, Deep Blue is a "cognitive wheel", and its programmers have no illusions that this is the case, but some observers have mistaken an exercise in parallel processing for an AI project. Is anybody claiming that playing chess against a computer is the same as playing a real person? Deep Blue's

[177] "Cognitive Wheels", p.146.
[178] Ibid, p.147.

programmers certainly aren't. The depth of the prejudice against machines is nicely dramatised in Isaac Asimov's story "A Boy's Best Friend". Jimmy Anderson lives on the Moon and has a robot dog called Robutt. One day his father brings him a real dog from Earth. "You won't need Robutt anymore" he announces "some other boy or girl will have Robutt." Jimmy is not impressed and objects.

> "What will the difference be between Robutt and the dog?"
> "It's hard to explain," said Mr Anderson, "but it will be easy to see. the dog will really love you. Robutt is just adjusted to act as though it loves you."
> "But, Dad, we don't know what's inside the dog, or what his feelings are. Maybe it's just acting, too."[179]

Jimmy's argument seems like a clincher, and in the context of AI it is. Unfortunately, his is precisely the argument that is used against non-human living creatures, like chimpanzees, who demonstrate some level of intelligence - maybe they are just simulating intelligence. Primate research is riven with such complaints. Doubts have been cast on research methods, and projects have foundered because research grants have expired. Many observers want primate language programs to fail, just as they want AI projects to fail. These observers want to preserve the illusion that human beings are unique, and do not feel comfortable being compared with chimpanzees or robots.

The Monkey in the Mirror

One of the most remarkable pieces of film footage I have ever seen involves an experiment involving monkeys and a mirror.[180] Monkeys, as opposed to apes, cannot recognise themselves in a mirror. They will challenge and attack the image as if it is another monkey. The chimpanzee in the film, on the other hand, recognises itself within minutes. At first it challenges the image and tries to play with it, but very quickly it realises that it is looking at itself. What I found remarkable is that the chimpanzee not only recognises itself after a few minutes, but rapidly begins to use the mirror to look at parts of itself which it cannot

[179] Isaac Asimov. "A Boy's Best Friend"
[180] *The Monkey in the Mirror, Natural World Series,* Produced by Karen Bass for the BBC Natural History Unit, Bristol, first broadcast February 1995. This film features footage of a number of the ape and researchers mentioned in this paper.

normally see, like the back of its ears. Apparently this ability does not arise in human children until they have reached 18 months.

Clearly there is not a gene in human beings and apes which enables them to recognise themselves in mirrors. There is however, a generalised fascination for seeing ones reflection - in water, as a shadow, in a photograph, in one's children. I have no idea how evolution could build such an ability through adaptation, but one has to conclude that something that is fundamental to human self-awareness is also operating in the chimpanzee as it looks in the mirror. On the other hand, I am wary of explanations which appeal to instinct and intuition to account for animal and human capacities.

Apparently it takes two years for a young seagull to learn what is edible and what is not. So, it spends the first two years of its life following older birds and literally taking the food from their mouths. This explains why seagulls are so aggressive, but it also serves to illustrate that much of the behaviour that we think of as instinctual is in fact learned.

Each Monday the cafe at the end of my street is circled by a wagon-train of all manner of wheeled devices for transporting children, as women with their babies converge on the place to talk about bringing up their infants. They talk about ways to hold the child, how to get it to sleep, feeding patterns, medical issues - why do they need to do all this? Because motherhood is not instinctual. Clearly we need to be careful about what we call a basic stimulus response mechanism.

It is difficult - perhaps even dangerous - to identify what is instinctive in animals and what is learned. My favourite example is the beaver. Beavers exhibit remarkable skill and resourcefulness as they shape their environment by building dams. The need to alter one's environment is a feature they share with human beings. However, in Wilsson's famous experiment[181] the beavers attempted to build a dam between two speakers emitting gurgling sounds. They may be able to change their external environment, but not their internal, mental environment.

Dennett considers the ability of human beings to alter their mental environment to be our most significant advance on other animals. In *Consciousness Explained* he argues that human beings are hard-wired to

respond to their environment in number of quite sophisticated ways. He suggests that language provides a level of virtual-wiring which makes us adaptable, so that unlike the beaver, we can often over-ride the hard-wired response.

If we follow Dennett's line of reasoning, what sets human beings apart from other animals (and robots) is precisely the fact that we are adaptable. When our world changes, we adapt. Dennett argues that consciousness arises from the conflict between various hard-wired and virtually-wired responses to a changing environment. This "pandemonium" theory of consciousness suggests that in any given situation the brain comes up with a plethora of conflicting responses and we act according to which one "wins". The feeling of consciousness is precisely that battle of competing solutions to the problem. Another way of looking at consciousness, Dennett suggests, is as a kind of narrative. We edit what we see of the world together into a narrative that more or less makes sense and more or less fits a world-view. According to this theory, if a robot is presented with a situation which conflicts with its model of its world, then we have the beginnings of consciousness. Nobody knows quite how this happens in human beings. Maybe we are looking for an extra cause when there isn't one to find. Maybe complex adaptive systems develop consciousness when they become genuinely complex. The possession of natural language by artificial intelligence seems to promise something that will make these machines far more than the sum of their parts. Language does not embody a model, but it is clear that it facilitates the building of mental models. As we have seen, the models need to be flexible enough to deal with unforeseen circumstances, and the AI mechanism must be able to adapt to changes, and alter its behaviour accordingly.

Carving the Turkey

Keeping track of changes in one's environment is a remarkable skill. Human beings have varied abilities in this area and so do animals. The fact that young children are so fascinated by disappearing tricks seems to suggest that keeping track of the changes is a skill well developed in young human beings. In computers, one needs to set up if/then/else loops to enable them to deal with

[181] L. Wilsson. "Observations and Experiments on the Ethology of the European Beaver" *Viltrevy, Swedish Wildlife* 8, 1974, pp.115-266.

changing situations, and the problem is that the number of ifs, thens, and elses in the real world is infinite. If our loop reads:

if turkey
then goto remove turkey routine
else goto find turkey routine

we can expect our robot to find the turkey in the fridge and remove it. Thus:-

pick up turkey
remove turkey from refrigerator
move to table
place turkey on table
return to main program

Under normal conditions the robot would proceed to carve the turkey. But suppose the turkey did explode when removed from the fridge? The robot would continue to the table and place the remains on the table and try to carve it. Consequently, we need to build in a series of subroutines allowing for mishaps concerning the turkey - exploding turkey, fake plastic turkey, very slippery turkey, live turkey etc. Perhaps the list isn't infinite, but the number of responses to each branch of the loop quickly becomes an exponential series. Human beings, and apes, have a way of breaking into exponential series by creatively using the unexpected or the error. It has been argued that being able to capitalise on error is the basis of creativity and the foundation of intelligence. Going back to our idea that being able to represent various futures to ourselves is crucial to consciousness and intelligence - consider again how one conceives such alternate plans.

If one removes a turkey from the fridge, but find that the table has been moved and the turkey hits the floor, the next time one removes a turkey from the fridge one will check the table position. One might even use the floor as the table, or even carve it in the fridge. Such creative responses require improvisation. Some improvisations won't work, but that is the nature of creative use of error. If it works you're a genius, if it doesn't you're an idiot. The first person to make a sandwich was a genius, subsequent sandwich makers don't always receive such acclaim.

Suppose our robot, Midnight, telephones the local deli and orders a sandwich - does this show intelligence? If we rebuild the kitchen from a television cookery

show and Midnight copies the movements it has already seen - would this count? Wouldn't it be just imitation? Where does imitation end and intelligent snack making begin?

There is a *3rd Rock from the Sun* episode in which the aliens try to understand what is so important about football. Sally, the alien military tactician, decides that the idea of the game is to "kill the guy who has the ball" and adds "If you make it to the big poles at the end you get to do a little dance."[182]This meditation on Earthlings and games highlights an important issue - a Martian or unschooled robot could never work out the rules of the game simply by watching it. It might be able to imitate all the moves, but without being instructed as to the goals and strategies it could not describe the game-rules. This is important because Midnight may be able to copy someone making the snack - and if lucky, make something edible - but put the robot in a kitchen it hasn't seen before, or vary any of the elements, and it would be unable to repeat its feat. Robots may be compared with actors - convincing when they stick to the script, but when they miss a cue the illusion falls to pieces. Every human activity has rules and strategies which enable a person to apply what it has learned in one situation to another. One needs to learn these rules and strategies, and language is the main means whereby we learn them.[183] Language, however, usually needs to be combined with other ways of teaching and learning. Imitation together with the acquisition of the rules and strategies through instruction and imprinting <u>is</u> learning.

To be considered intelligent, an agent must make decisions and overcome problems. Midnight must ask, and answer for itself, a lot of questions. How thick does one slice the bread? How much butter does one spread? How much mayonnaise? What is the correct way to pour beer? At any of these hurdles, Midnight could fall. Converting what it knows about snack-making, into useful action is to effectively make a link between two different ways of thinking.

[182] 3rd Rock From the Sun. Written by Michael Glouberman and Andrew Orstein .1996.

[183] It is the lack of genuine strategic ability which excludes Deep Blue from the ranks of the grand-master. If it could come up with a new chess strategy it might make the chess journals.

Knowing How and Knowing That

Hard-wiring provides a level of knowing "how to" do something which is independent of any knowing "that" something is the case. It is language that encourages us to look at "knowing how" as "knowing that". It enables us to stand back and look at what we are doing. This can sometimes be a disadvantage.

The hand eye co-ordination required in juggling seems to require a sophisticated knowledge of gravity. In fact, knowing how to juggle does not entail the juggler working out that "the time a ball spends in flight is proportional to the square root of the height of the throw." Nevertheless, that is the formula the juggler is "applying". Ethologists such as Dawkins have encouraged us to think about abilities such as hand-eye co-ordination as a product of our evolution. They suggest that our ape ancestors evolved a range of abilities in order to forage in the forest. Dennett himself points out that such skills have been inherited by human beings, but may now be employed to do very different things from those for which they were originally evolved. His point is that motor skills, for example, may be deep rooted hard-wired parts of the human organism.

When we pay attention to things like walking down stairs and catching a ball - that is, when we look at ourselves in the act of doing these things - we often become disorientated and the impossibility of the act overwhelms us. Looked at "objectively", these abilities seem overwhelmingly complex. Using our language-given abilities to stand back and look at what we are doing can actually be a handicap. It is therefore not surprising that teaching such physical activities to robots has proven staggeringly difficult. Robots that walk on two legs without falling over unexpected obstacles are still a rarity.[184] A computer can calculate the trajectory of a 10 year trip to Mars, but calculating the trajectory of its feet going down stairs - that's a different matter. When Brooks says that some analog aspects of robot activity cannot be digitalised, he is declaring that a robot which can act intelligently in the world is the embodiment of two kinds of intelligence - roughly corresponding to mental and physical. Furthermore, he argues that "classical and neo-classical AI" has made a fundamental error.

> both approaches make the mistake of assuming that because a description of reasoning/behavior/learning is possible at some level, then that description must be made explicit and internal to any system that

[184] I understand that in 1997 a Japanese lab developed a stair climbing robot. I have been unable to find a paper relating to how it works.

> carries out the reasoning/behavior/learning. This introspective confusion between surface observations and deep structure has led AI away from its original goals of building complex, versatile, intelligent systems and towards the construction of systems capable of performing only within limited problem domains and in extremely constrained environmental conditions.[185]

Computer metaphors of how human beings think encourage the view that the really intelligent and advanced part of us is the "software" - which in Dennett's account is language and our ability to represent things to ourselves in our minds. The hard-wired physical stuff is possible in principle to embody in a robot but not so easy in practice. Perhaps the ability to catch a ball or balance on a tightrope are the really advanced things - certainly they are complex. Perhaps equipping computers with massive digital computing power is putting the cart before the horse? Theoretical physics requires a lot of "software", but playing the piano requires the "software" plus physical skills. Brooks argues that being embodied is a crucial aspect of human intelligence. His approach is try to solve the physical problems of embodiment using engineering as opposed to digitalisation.

> For example, when putting a jug of milk in the refrigerator, you can exploit the pendulum action of your arm to move the milk (Greene 1982). The swing of the jug does not need to be explicitly planned or controlled, since it is the natural behavior of the system. Instead of having to plan the whole motion, the system only has to modulate, guide and correct the natural dynamics. For an embodied system, internal representations can be ultimately grounded in sensory-motor interactions with the world.[186]

There is a prevalent myth that the information age is going to bring about a revolution in how we think and what we can do with our minds. Scientists such as John Barrow and Frank Tipler have speculated on something called the Omega point, which is reached when life expands to fill the entire universe.

> At the instant the Omega point is reached life will have gained control of all matter and forces not only in a single universe, but in all universes whose existence is logically possible; life will have spread into all spatial regions in all universes which could logically exist, and will have stored an infinite amount of information including all bits of knowledge which it is logically possible to know. And this is the end.[187]

---

[185] "Alternate Essences of Intelligence"
[186] Ibid.
[187] John D. Barrow and Frank J.Tipler. *The Anthropic Cosmological Principle*. Oxford: Oxford University Press, p.677.

This kind of nonsense is favoured by mathematicians who have calculated how many particles there are in the universe and how many interactions there could possibly be until we all disappear in a kind of entropic whimper. These people are making the mistake of assuming that because one can describe a process at one particular level, they have described what is going on at all levels.

The frame problem exposes these speculations for the nonsense they are. The universe isn't just a bunch of facts to be stored as information. Fact are only facts when they are relevant facts, and such facts are not truths, or even untruths. Dawkins and Dennett might describe them as memes, but even if you call them ideas, there is no limit to them. Even if you add up all the particles in the universe and multiply them by the number of possible time intervals, multiplied by the possible universes, and then imagine all the opposite cases - this exercise says nothing about knowledge, ideas or the limits of the human mind.

Someone's Got Some Explaining To Do

The ideas which arise in mind of man are, according to Dawkins, like the dams of beavers or the nests of birds - extended phenotypes. Like feathers, or fur, they are manifestations of the organism, but they extend into the environment. In the case of human beings, the dams are intricately woven selves consisting of knowledge, beliefs and practices which we have inherited - partly genetically - but largely culturally, from our forebears.[188] Imagine a world where beavers could pass on improvements in dam design from generation to generation - pretty soon beaver dams would be elaborate structures - and if beavers were clever enough, the dams would have sophisticated water-level management systems etc.. "Dam design" amongst human beings has developed rapidly over the last 10,000 years due to enhanced communication abilities. Enhanced communication skills are necessary to promote social organisation, and essential in order to pass on the crucial social and tool-making skills necessary to survive in a complex social system and a changing environment.

For Chomsky, the great problem with infant development is the leap from stimulus-response use of language to symbolic language. For anthropologists, the problem is explaining how human beings made the leap from ape-like creatures to what we are now. What caused us to evolve this large sophisticated

[188] See Consciousness Explained for Dennett's account of how we weave a consciousness.

brain? What environmental pressures could cause such a radical change? What made our ape -like ancestors shed their fur, adopt an upright position and start talking? Despite its great explanatory power, neither science nor Darwin has been able to provide a solution. The most prevalent explanations imagine that our ape-ancestors were somehow isolated in an environment which was different from our close ape relatives - forced onto the savannah, or marooned on an island. None of these accounts convincingly model the development of human intelligence and language. No wonder science fiction writers, such as Arthur C. Clarke imagine our species being helped along in our development by god-like aliens. In Mary Shelley's *Frankenstein*, it is lightning which brings the monster alive. In *Short Circuit,* Number 5 becomes sentient after a lightning strike causes it to malfunction. Number 5 is the dramatisation of the view that sentience is a malfunction.

In the absence of creative malfunction, AI researchers are hoping that language holds the key to intelligent modelling, just as it seems to be the decisive factor in human intelligence. In fact, we may find that the possession of language pre-supposes that the frame problem is largely solved and that looking for an answer to the frame problem in language is like using the solution to get to the problem. Solving the frame problem is a pre-requisite for solving the language problem.

The frame problem in AI cannot be solved by giving a robot thousands of common sense facts about the world. Being an agent in the world and reacting to things in the world has its own "grammar". There is a kind of physical logic to moving around in space, moving things around, and reacting to stimuli. This logic is not programmable. Brooks comments on the introductory page of the MIT AI Lab web-site.

> There has been a realization amongst many people at our Lab that the keys to intelligence are self adapting perceptual systems, motor systems, and language related modules. This is in marked contrast to earlier approaches that focused on reasoning, planning, and knowledge representations as the keys to Artificial Intelligence.[189]

I am inclined to agree with Brooks that systems which adapt will be more able to develop useful language skills. The approach at Brooks' lab emphasises

[189] Rodney A. Brooks (brooks@ai.mit.edu)

metaphors of childhood training, and more and more examines what factors are important in the development of a human child. Brooks believes that the essences of human intelligence lie in embodiment, social interaction, and integration with the environment.

Natural language has to be learned in conjunction with being embodied in the world and reacting to it. The brain in a vat could not learn language, and it is impossible to download natural language into a computer. CYC is a computer-brain in a vat, and will never develop natural language. Cog on the other hand has its foot (if it has a foot) on the first rung of the ladder to language. It is dealing with the hard, frame problems from the bottom up, and will learn language only after it has largely overcome these problems of being an agent in the world.

Appendix A: The Dance Language and Orientation of Bees

Excerpt from Karl von Frisch, 1967. The Belknap Press of Harvard University Press, Cambridge, Mass. pp. 4-5.

When I wished to start an experiment I would set out on a table in the open a sheet of cardboard with some honey on it. As a rule this was found after a few hours by one bee. Then their number would grow swiftly to dozens, or even to hundreds. A further phenomenon too spoke emphatically in favor of communication among the hivemates. I trained bees to collect from a dish filled with sugar water. When the feeding dish had been emptied I paused in order to limit the number of collectors. At first large numbers of bees swarmed about the empty dish, but gradually they dispersed and after about 20 min they visited only sporadically. But if one of these now found the dish refilled, the others reappeared in swift succession after her return home.

I was curious to know how the news was spread about there at home, and built an observation hive with glass windows...In the spring of 1919 I set it up..., erected a feeding station beside it, and marked the forager bees with a spot of red pigment on the thorax. After a pause in the offering of food, they would sit among the others on the comb near the hive entrance.

The next scout found the dish refilled. It was a fascinating spectacle when after her return home she performed a round dance, in which the red -spotted bees sitting nearby showed lively interest. They tripped along after the dancer, and then left the hive to hasten to the feeding station. Soon it became apparent that the circular running is a dance of invitation, which not only recalls the former collecting group to action but also recruits new members to strengthen the working party... With pollen collectors that were returning home with filled pollen baskets from natural sources of provisions, I saw another form of dance, the tail-wagging dance, and fell into the error of thinking that the round dance was performed when sugar water or nectar was collected and the tail-wagging dance after pollen collecting (v. Frisch 1923). Henkel (1938) refuted this. Under natural conditions he observed with nectar collectors tail-wagging dances that did not

differ in form from pollen collectors, and explained the round dances of the sugar -water collectors as due merely to the unnatural abundance of food at my artificial feeding stations. On the basis of new experiments I at first held to my conception (1942). Today we know that Henkel was right when he described tail-wagging dances performed by collectors of nectar, but I was right too in describing their performance of round dances. I was wrong when I regarded the round dances as dependent on the gathering of nectar, and he was wrong in ascribing them to the abundance of food. The clarification came when I gave my co-worker Ruth Beutler a piece of bad advice. She was running a feeding station with the odor of thyme 500 m away from a beehive and wanted to have the bees gather quickly around a sugar-water dish at a place nearer the hive. I advised her to feed them well at the 500-m station and also put out a sugar-water dish with thyme fragrance at the desired place near the hive. The hivemates would be stimulated by the round dances of bees harvesting from the distant point to search first nearby around the hive and would necessarily find the new feeding dish quickly. There was no success. Did the distance of the feeding place influence the manner of dancing?

Experiments directed to this point showed in fact that round dances were performed with sources of food nearby, tail-wagging dances with more distant ones, by collectors of nectar just as by collectors of pollen, and that the tail -wagging dances announced also the direction and distance of the goal. The mistake had come from my setting up the artificial feeding station with its sugar water in the immediate neighborhood of the hive, in order to keep both feeding station and comb in view, whereas the pollen collectors were coming from natural, more distant sources. Under these conditions there were only round dances among the bees collecting sugar water, and only tail wagging dances among the pollen collectors...Probably bad advice has rarely been so nobly rewarded."

Appendix B: KB interchange standards

Doug Lenat <lenat@mcc.com>

* Mail folder: Interlingua Mail

* Next message: sowa@watson.ibm.com: "Tools to Enable Knowledge Sharing"
* Previous message: Tracy Schwartz: "KB interchange standards "
* Maybe in reply to: Tracy Schwartz: "KB interchange standards "
* Reply: Robert Neches: "Lenat's note (was Re: KB interchange standards)"

Date: Wed, 27 Nov 1991 12:56-0800
From: Doug Lenat <lenat@mcc.com>
Subject: KB interchange standards
To: interlingua@isi.edu, kr-advisory@isi.edu, SRKB@isi.edu, krd@ai.mit.edu, james@cs.rochester.edu, davis@ai.mit.edu, feigenbaum@sumex-aim.stanford.edu, forbus@ils.nwu.edu, rkahn@nri.reston.va.us, pkarp@ai.sri.com, kunz@intellicorp.com, jin@eagle.mit.edu, luu@isi.edu, malone@eagle.mit.edu, overt@prc.unisys.com, porter@cs.utexas.edu, dan_russell.parc@xerox.com, bwilliam@parc.xerox.com, hewitt-srkb@ai.mit.edu, mars@cs.utwente.nl, cleary@corwin.ccs.northeastern.edu, doug@csi.uottawa.ca, john@atc.boeing.com, roger@ci.deere.com, gio@darpa.mil, friedland@ptolemy.arc.nasa.gov
Cc: lenat@mcc.com, guha@mcc.com
Message-id: <19911127205633.3.DOUG@SURYA.CYC-WEST.MCC.COM>

We think that the time may be right, now, for this sort of push on knowledge-sharing. In some ways, our experiences with Cyc can serve as a microcosm for this sort of inter-group interaction. One interesting result is that many of the problems you're talking about still remain, even when the cooperating groups all use exactly the same representation system (syntax, vocabulary/terms, and inference machinery.)

Let us explain that "microcosm" remark. Superficially, this occurs because we have many different groups using and helping to build Cyc, located around the country. The analogy holds at a deeper level as well, even within a single site: Our knowledge enterers work in small teams, often for weeks at a time, building up a "micro-theory" of some topic. They must have some level of interaction with other groups, and already-entered micro-theories, but the less interaction the faster they can work.

Much of the attention and controversy of the Standards for KB Interchange effort seems to focus on sharing syntax and semantics of a language, and a little about sharing vocabulary. Well, let's consider Cyc's knowledge enterer teams, since they do share these things. Does it solve the problem? If not, what else is/was needed?
One of the recurring problems during 1984-1989 was "divergence" --- DESPITE the aforementioned sharing. Different groups would use a term slightly differently in their new micro-theory (compared to the way it had been used before in other theories, sometimes even by themselves at an earlier time.)

The standard solution to this would be to pick a small set of primitives, and lock in their meanings. The problems in our case -- and yours -- are (a) there is no small set, and (b) it's almost impossible to nail down the meaning of most interesting terms, because of the inherent ambiguity in whatever set of terms are "primitive."

So what did we do?

(1) For one thing, we insist only on local coherence. I.e., groups share most of the meaning of most of the terms with other groups, but within a group (working on a particular micro-theory) they strive for complete sharing.

(2) For another thing, both kinds of sharing are greatly facilitated by the existing KB content --- i.e., if the terms involved are already used in many existing axioms.

While (2) can be achieved through massive straightforward effort, (1) is more subtle, and has required certain significant extensions to the representation framework. More specifically, we had to introduce the whole machinery of contexts/micro-theories into Cyc (which is why "divergence" has been much less of a problem, since 1990.)

Each group enters its micro-theory into a context. Different contexts may use different vocabularies, may make different assumptions, may contradict assertions made in other contexts, etc. (Each context is a first class object in our language, and instead of saying that a formula is either universally true or false, it can be true in some contexts and false (or even unstatable) in others.)

Both knowledge entering and problem solving go on in a context. Axioms external to a context are imported (lifted) from other contexts, using articulation rules. So the question of `what to share' is partially decided at knowledge-entering time, by humans, and partially at inference time, by the system.

>From this, it seems that an optimal knowledge-sharing effort should attempt to build on a significant (large, broad) existing base KB, and it should incorporate some sort of context mechanism, so that the sharing can be flexible and, if necessary, reasoned about by the system.
If there is sufficient call for it, we'd like to try to find some way to share Cyc -- its content and context mechanism, as well as the less-important syntax and vocabulary of its language -- with you. Think of it either as a seed, or as scaffolding, but in any case we feel that something like it (in both breadth and size, which is currently over a million axioms) is going to be needed to serve as the semantic glue to enable the sort of knowledge sharing we all have in mind.

Sincerely,

Doug Lenat and R. V. Guha

Bibliography

Books

Adams, D. & M. Carwardine. *Last Chance To See*. Heinemann: London, 1990.

Adams, D. *So Long, and Thanks for All the Fish.* London: Pan Books, 1984.

Appleyard, B. *Understanding the Present.* London: Picador, 1992.

Asimov, I. “A Boy’s Best Friend”

_____. "The Bicentennial Man." rpt. in *Machines That Think.* ed., Isaac Asimov, Patricia Warrick, and Martin Greenberg. Harmondsworth: Penguin, 1986.

Asimov,I., Patricia S.Warwick, and Martin H.Greenberg, eds. *Machines That Think*. Harmondsworth: Penguin, 1983.

Barrow, J.D. and F.J.Tipler. *The Anthropic Cosmological Principle*. Oxford: Oxford University Press.

Barthes, R. "Inaugural Lecture to College de France” (1977) in *A Barthes Reader*.

Bell, J.S. *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press, 1987.

Benford, G. “Alphas” collected in *Best New SF4* edited by Gardner Dozios. London: Robinson, 1990.

Brecht, B. *The Life of Galileo.* translated by John Willett, London: Methuen, 1980.

Brooks, R.A. “From Earwigs to Humans” MIT Artificial Intelligence Laboratory, 1996 (see Web Sites)

Brooks, R.A. and Lynn Andrea Stein. *Building Brains for Bodies*. MIT Artificial Intelligence Laboratory Memo 1439, August 1993.

Brooks, R.A. and Cynthia Breazeal (Ferrell), Robert Irie, Charles C. Kemp, Matthew Marjanovic, Brian Scassellati, Matthew Williamson. *Alternate Essences of Intelligence* (Submitted to AAAI-98), January 1998.

Brown, F. “Placet is a Crazy Place”

Carroll, J.B. *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf*. The M.I.T. Press, 1956

Chomsky, N. *Problems of Knowledge and Freedom*. London: Fontana, 1972.

_____. *Language and Problems of Knowledge: The Managua Lectures*. Massachusetts: The MIT Press, 1988.

Cytowic, R.E. *The Man Who Tasted Shapes.* London: Abacus, 1994

Dawkins, R. "The Moon is not a Calabash" in *The Times Higher Educational Supplement*, 30th September 1994.

_____. *The Extended Phenotype.* Oxford & San Francisco: Freeman, 1982.

_____. *The Selfish Gene.* Oxford: Oxford University Press, 1976.

Delany, S. *Babel-17*. rpt. London: Sphere, 1966.

Dennett, D.C. *Darwin’s Dangerous Idea*. Harmondsworth: Penguin, 1996.

_____. “Cognitive wheels: the frame problem of AI", *Minds, Machines and Evolution,* edited by Christopher Hookway, Cambridge University Press, 1984.

_____. *Consciousness Explained*. Harmondsworth: Penguin, 1991.

Derrida, J. "Structure, Sign and Play, in the Discourse of the Human Sciences" in *Modern Literary Theory: A Reader* (2nd Ed) edited by Philip Rice and Patricia Waugh, London: Edward Arnold, 1992.
Dick, P.K. "The Electric Ant." *The Magazine of Fantasy and Science Fiction*, 1969. Reprinted in *Machines That Think*. Asimov, Warrick, Greenberg eds., 1986.
_____. *Ubik.* London: Granada, 1975.
Dickson, G.R. "The Monkey Wrench" reprinted in *The Penguin Science Fiction Omnibus*, Harmondsworth: Penguin, 1973.
Ellis, J.M. *Against Deconstruction*. New Jersey: Princeton University Press, 1989.
Freedman, D. H. “Commonsense and the Computer” *Discover*, August 1990, pp. 65-71.
Gell-Man, M. *The Quark and the Jaguar*. London: Abacus, 1995.
Gödel, K. *On Formally Undecidable Propositions*, translated by J. van Heijenoort, in *From Frege to Gödel: A Source Book on Mathematical Logic*, 1879-1931, ed. by J. Van Heijenoort. Cambridge, Mass: Harvard University Press, 1967.
Goodwin, B. *The Times Higher Educational Supplement*, September 30 1994.
Gould, S.J. on Charles Darwin's *Origin of Species* in *The Times Higher Educational Supplement*, September 30 1994.
Gould. "The Evolution of Life on the Earth" in *Scientific American: Special Issue, Life in the Universe.* October 1994.
Gribbin, J. *Schrodinger's Kittens and the Search for Reality.* London: Little, Brown and Company, 1995.
Hockett, C.F. *The View From Language: Selected Essays from 1948-1974*. Athens: University of Georgia Press, 1977.
Holland, J.H. “Genetic Algorithms” in *Scientific American*, July 1992.
Jakobson, R. and M. Halle, *The Fundamentals of Language*, The Hague: Mouton, 1956.
Jameson, F. *The Prison House of Language: A Critical Account of Structuralism and Russian Formalism*, Princeton, 1972
Janlert, Lars-Erik. “Modelling Change: The Frame Problem”, in *The Robot’s Dilemma: The Frame Problem in Artificial Intelligence* edited by Zenon W. Pylyshyn, New Jersey: Ablex Publishing Corporation, 1987.
_____. “The Frame Problem: Freedom or Stability? With Pictures We Can Have Both.” in *The Robot’s Dilemma Revisited*. edited by Kenneth M. Ford and Zenon W Pylyshyn. New Jersey: Ablex Publishing Corporation, 1996.
Lenat, D.B. “Artificial Intelligence” in *Scientific American*. September 1995.
McCrone, J. *The Ape That Spoke: Language and the Evolution of the Human Mind.* London: Picador, 1990.
Omni, December 1994 article on Richard Hoagland.
Orwell, G. *Nineteen Eighty-Four,* Harmondsworth: Penguin, 1949
_____. "Politics and the English Language" in *The Collected Essays, Journalism and Letters of George Orwell*. Ed. Sonia Orwell and Ian Angus. 4 vols. London: Secker and Warburg, 1968.
Pinker, S. *The Language Instinct*. New York: William Morrow, 1994.

Roberts, A. *Arena Magazine.* Feb-March 1993, pp.34-36. The article is based on a paper entitled "Interventions" delivered at the Monash Craft Conference in 1992.
Roget's Thesaurus. Abridged by Susan M. Lloyd, Harmondsworth: Penguin, 1984.
Searle, J. "Is the Brain's Mind a Computer Program?" in *Artificial Intelligence: A Debate.* in *Scientific American,* January 1990.
Shulyer, W.M., Jr. "Could Anyone Here Speak Babel-17?", in *Philosophers Look at Science Fiction* edited by Nicholas D. Smith, Chicago: Nelson-Hall, 1982.
Silverberg, R. "The Macauley Circuit" 1956. Reprinted in *Machines That Think* edited by Isaac Asimov, Patricia S.Warwick, and Martin H.Greenberg, Penguin, 1983
Bullock, A. & O. Stallybrass, eds. *The Fontana Dictionary of Modern Thought.* London: Fontana, 1977.
Tennekes, H. *The Simple Science of Flight: From Insects to Jumbo Jets*, MIT Press, 1996.
Time Magazine, April 1, 1996, includes a debate entitled *Can Machines Think?*
van Vogt, A.E. *The Quest for the Future*. London: NEL, 1972.
Webb, B. "A Cricket Robot", *Scientific American,* December 1996.
Weizenbaum, J. *Computer Power and Human Reason*. (2nd edition), Harmondsworth: Penguin, 1984.
Whorf, B.L. "An American Indian Model of the Universe" in *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf* edited by John B. Carroll. MIT Press: Massachusetts, 1956.
Wilsson, L. "Observations and Experiments on the Ethology of the European Beaver" *Viltrevy, Swedish Wildlife* 8, 1974, pp.115-266.
Wittgenstein, L. *Philosophical Investigations*. Oxford: Basil Blackwell, reprinted 1967.

Bibliography Continued

Films

Dark Star directed by John Carpenter and written by John Carpenter and Dan O'Bannon, 1974.

Independence Day, 1996.

The Monkey in the Mirror, Natural World Series, Produced by Karen Bass for the BBC Natural History Unit, Bristol, first broadcast February 1995.

Short Circuit directed by John Badham, 1986.

Television Series

3rd Rock From the Sun. Written by Michael Glouberman and Andrew Orstein .1996.

Red Dwarf Episode "Backwards".

***Star Trek: The Next Generation* - Paramount Pictures**

"Elementary My Dear Data" written by Brian Alan Lane, 1988.

"The Measure of a Man" written by Melinda M.Snodgrass and directed by Robert Sheerer, 1989.

"Data's Day", teleplay by Harold Aptor and Ronald D. Moore, 1990.

"Peak Performance", written by David Kemper, 1990.

"The Outrageous Okona", teleplay by Harold Aptor and Ronald D. Moore, 1990.

Web Sites

brooks@ai.mit.edu, 1996 for a review of progress on Cog.

Cybercinema. http://www.english.uiuc.edu/cybercinema/main.htm

Journal.Graphics@bobj.cc.uic.edu Newsgroups: jrnl.pbs.nova Subject: [1/5] Can Chimps Talk? Date: Sat, 19 Feb 1994

Lenat, D.B. Memo from Doug Lenat via Interlingua Mail, 27th Nov 1991.See Appendix B.

Scientific American Commentary on Kasparov vs Deep Blue on May 4th 1997. http://www.sciam.com/explorations/042197chess/050597/chesscom.html

Whitten, D. "The Unofficial, Unauthorised CYC Frequently Asked Questions Information Sheet."