UNIVERSITY OF NAVARRA

ECCLESIASTICAL FACULTY OF PHILOSOPHY

Mark Telford Georges

THE PROBLEM OF STORING COMMON SENSE IN ARTIFICIAL INTELLIGENCE

Context in CYC

Doctoral dissertation directed by

Dr. Jaime Nubiola

Pamplona, 2000

Table of Contents

Abbreviations

AAAI	American Association for Artificial Intelligence
ACM	Association of Computing Machines
AI	Artificial Intelligence
ASR	Automatic Speach Recognition
CBR	Case Based Reasoning
CYC	The name comes from en-CYC-lopedia
DCS	D. B. Lenat, "The Dimensions of Context Space", Austin, TX: Cycorp, 1998.
EH	The Empirical Inquiry Hypothesis
KB	Knowledge Base
NLU	Natural Language Understanding
OTK	D. B. Lenat and E. A. Feigenbaum, 1991, "On the Thresholds of Knowledge", *Artificial Intelligence* 47: 185-250.
PDP	Parallel Distributed Processing

Introduction

Does computer science have anything to do with philosophy? As a computer scientist myself I can testify that my colleagues generally regard philosophy as distant and confused with little relevance to their professional vocation as computer scientists. In Artificial Intelligence (AI), there are some signs of an opening up to philosophy and vice versa. Nevertheless, the literature on AI generally maintains a distant attitude. Douglas Lenat, for example, a leading innovator in AI, has written extensively on what he calls "ontological engineering". In his book *Building Large Knowledge Based Systems*, he is careful to clarify that such ontology is: 'empirical, experimental engineering, as contrasted with ontological theorizing, which philosophers have done for millennia.'[1]

This dissertation is a testimony that computer science and philosophy are not distant at all because it is the same human person that studies computers and philosophy and both study one and the same reality though perhaps from different perspectives. Philosophers seek to better understand the anthropological and ontological principles that characterize and govern behavior. Such work permits a clearer appreciation of the limits and scope of AI and provides orientation for the construction and evaluation of computational models of the mind.

The goal of this investigation is to take time out on behalf of the computer science community, in particular on behalf of those that I have worked more closely with over the years, to reflect on the path followed by AI and to examine where we may have strayed from the deeper motivations, desires, and objective limits that ought to configure the way for exciting and fruitful AI research. For this exercise, it is necessary to consider concrete cases, seeking out and evaluating the underlying motivations and tendencies. The CYC project to store common sense gathers together the experience of many years of AI research and is widely considered a forefront project in AI. It is an ideal candidate for the exercise that this work proposes and for this reason this investigation has centered around the CYC project, led by Douglas B. Lenat.

---

[1] D. B. Lenat and R. V. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project*, Reading, MA: Addison-Wesley, 1990, p. 23.

AI researchers have recognized the need to incorporate knowledge into their programs in order to address the deeper issues of general intelligence. Programs in the early days of AI relied heavily on compiled knowledge and formal methods for problem solving. They were highly competent in certain limited domains, but entirely useless in others. An important lesson learned over the years is that general intelligence cannot be separated from general knowledge. Knowledge is needed to communicate, knowledge is needed to reason, knowledge is needed to learn.[2] From the point of view of cognitive science it can be said that the experiments realized in AI to date have invalidated theories and models of intelligence that rely heavily on formal methods and little on knowledge.

Douglas Lenat has seriously taken up the challenge to build a system that is knowledge based. He and his colleagues at CYC have been hard at work for over fifteen year on a three phase project to harness and manipulate the day to day facts of common sense knowledge. Building CYC has been challenging and arduous. The representation scheme and the inference mechanisms that CYC uses has been reworked several times to overcome conceptual and technical shortcomings. Lenat and his team at CYC originally used a simple frame and slot representation language to input and relate knowledge items in CYC. The language was not flexible enough to express peculiarly pragmatic phenomenon such as contexts, opinions, expectations, counterfactual conditionals, and other similar realities in an efficient manner. Much of the knowledge that we ordinarily use proved too complex to for the frame and slot representation scheme. It was quickly abandoned in favor of second order predicate calculus which caters for uncertainty and beliefs.

The difficulties encountered in harnessing common sense has given rise to what Lenat considers the most important lesson learnt - the need to incorporate contexts. Lenat admits that it was foolhardy to try to maintain consistency in one huge flat CYC knowledge base. The CYC KB has been divided up into hundreds of contexts or microtheories. Each context is consistent within itself, but there can be contradictions among them. On the inference front, a method of argumentation has been incorporated. To decide whether to believe something, CYC gathers up all the pro and con arguments it can think of, examines them, and then reaches a conclusion.

While microtheory technology promised to overcome the early difficulties, important shortcomings arose which undermined and offset its usefulness. The copy and edit mechanisms to create and modify contexts turned out to be a bothersome source of logical inconsistencies. Furthermore, as more contexts were created it became difficult and time consuming for the knowledge enterer to select the most adequate context for new assertions. Finally, Lenat and his team were faced with the difficult challenge of maintaining a good intermediate value for the number and choice of contexts. Though microtheory technology provided ways to overcome early difficulties, the shortcoming

---

[2] Cfr. D. B. Lenat and E. A. Feigenbaum, "On the Thresholds of Knowledge", *Artificial Intelligence* 47 (1991), p. 196.

encountered during its implementation suggested the need for new theories and techniques.

Lenat's new strategy addresses the shortcomings of microtheory technology. He proposes that context can be adequately defined by specifying values along 12 independent dimensions. Proceeding in this way, Lenat upholds that the difficulties encountered in the previous strategy can be overcome.

A great deal of study and experimentation has been done at CYC to better understand contexts, to determine how contexts can be represented and how a program can reason using contexts. The advances and obstacles that Lenat has encountered serves to better understand the philosophical influences in the project and to evaluate Lenat's goals and tendencies. For this reason I have chosen to center this investigation around context in CYC.

In the first chapter, a brief sketch of the history of artificial intelligence is offered. The context in which the problem of storing common sense arose is explained, giving special attention to the underlying philosophical issues.

Lenat considers AI an empirical inquiry that investigates real intelligence and it is in this context that his work is best understood. In the second chapter I thus describe CYC from a scientific perspective, that is, as a scientific experiment. The methodology of AI to which Lenat subscribes is presented and the theories that CYC tests are described. In the fourth section, CYC's ontological scheme is described and in the final section the history of the project is presented.

In the third chapter a detailed description of the work that has been done to represent contexts in CYC is offered. How the models and theories used have evolved is explained and a detailed description of Lenat's new strategy is presented.

The theories and models that have been adopted in CYC to represent and reason using contexts are fruit of a particular way of thinking. In chapter four the underlying principles of the CYC project are discussed and the strengths and weaknesses of the project are evaluated. The first part of the discussion concerns the methodology that is adopted in CYC. In the second part of the chapter the principles, strengths and weakness of CYC theory are discussed and in the third section I analyze and evaluate the efforts that have been made to store context. In the fourth section, the role of philosophical evaluation in AI is discussed, an activity that Lenat and his colleagues purposely avoid. The dissertation is closed with a brief conclusion.

At this point I wish to thank all those who have been a constant point of support in the arduous elaboration of this thesis. I wish to thank especially my director Jaime Nubiola for his generous attention and Manuel García Clavel who suggested the topic for this research. I extend my thanks to Mariano Artigas for his cordial support and stimulating conversation. His thought has inspired many of the ideas herein exposed. I wish to express my gratitude to my colleagues in the philosophy department at the University of Navarra, especially those in the Peircean studies group. Our frank and lively conversations have been a continual source of inspiration and motivation. I thank

in a special way the Adveniat foundation for their economic assistance. Due to their generous aid I have been able to meet the financial demands of my studies. Last but not least, I wish to express my deepest feelings of solidarity to my colleagues in the computer science profession. It is as a service to them that I embarked on this investigation.

Chapter I

A short history of Artificial Intelligence

The history of artificial intelligence is part of an ongoing drama, with man as protagonist, to conquer the enigmas of his own self and of the world around him. If we can simulate man, or even better, replicate man, then there is no more mystery. We would have come to the closing act of the drama where science and technology definitively displace mystery. In the meantime, however, the drama continues. It is a soul-wrenching story. On one hand, through the toil of our hands, we have spectacular exhibits in AI such as Deep Blue, the world's no. 1 chess player, and several speech recognition programs are already in the marketplace. At the same time, however, as we seek to replicate intelligence, the enigma of man seems more imposing, more overwhelming. Most of the experts agree that generally intelligent machines such as *HAL* (the intelligent computer in *2001: A Space Odyssey*) are far off if not simply illusory.[3]

In the first section of this chapter, I briefly sketch the history of AI and discuss several uses of the term "Artificial Intelligence". In the second section the principle trends in AI are explained. In the third section the underlying issues in the philosophy of AI are outlined and a few representative arguments are presented. In the fourth section I describe how AI has fared, illustrating how the concern for knowledge in AI has taken root. In the final section I consider various aspects of the problem of storing common sense and introduce the CYC project led by Douglas B. Lenat.

1.1. The origin and use of the term "Artificial Intelligence"

The field of artificial intelligence was given its name by John McCarthy, one of the landmark pioneers in the field. His pioneering work includes the introduction of LISP at MIT in 1958, which is the computer language that is used for the vast majority of AI programs. In 1956, McCarthy organized the conference that AI researchers consider as marking the birth of their subject.[4] McCarthy wanted to bring together the people he

---

[3] R. Schank, "How Could HAL Use Language" in D. G. Stork, ed., *Hal's Legacy: 2001's Computer as Dream and Reality*, Cambridge, MA: The MIT Press, 1997, p. 189.

[4] Cfr. E. Charniak and D. McDermott, *Introduction to Artificial Intelligence*, Reading, MA: Addison-Wesley, 1985, pp. 9-11.

knew who were keen on making intelligent machines. He invited them to a two month long summer workshop of brainstorming at Dartmouth College in New Hampshire. The conference was entitled: *The Dartmouth Summer Research Project on Artificial Intelligence*. The conference ended but the last two words remained as the logo for a new and multiform discipline to emulate the human mind.

The Dartmouth conference was the first meeting place of the four men who would lead artificial intelligence, at least in the United States, for the next twenty years: J. McCarthy, M. Minsky, A. Newell and H. Simon. Despite the organizational efforts, however, relatively little was accomplished at the conference. Most of the people that McCarthy invited came and went haphazardly throughout the two month long endeavor. Despite this however, the conference had a synergetic effect. What had previously been a scattering of individual enthusiasts working in relative isolation was suddenly a scientific community with its own research goals and a strong sense of self-identity. In the years following Dartmouth, artificial intelligence laboratories were established at a number of universities. In 1959 Minsky and McCarthy established an AI lab at MIT. Three years later, McCarthy moved to Stanford to found an AI laboratory there leaving Minsky at the helm of the MIT AI laboratory. AI labs were later installed at Carnegie Mellon under Newell and Simon and at Edinburgh under Donald Michie, a leading figure of the British AI scene.[5]

1.1.1. Influences in AI

AI grew out of extremely complicated circumstances in which technical advances, philosophical currents, popular culture and commercial interests all have a part to play.[6] Descartes, Darwin and Kant have a lot to do with Isaac Asimov, HAL, the pentium chip, LISP, computer prices and economic health when it comes to making sense out of artificial intelligence. Furthermore, the great deal of research in the field in recent years has given rise to numerous ramifications in the field generally understood as Artificial Intelligence. While one AI lab, for example, may be dedicated to speech recognition, others may focus on computer vision, human interface design or programming languages, to name just a few of the divisions that have arisen.

Given the complex interplay of technical advances, philosophical currents, popular culture and commercial interests in the forging of AI, and the many ramifications in the field, it is not surprising that the term "Artificial Intelligence" can be used in many different senses. Douglas Lenat subscribes to a particular interpretation which will be considered in more detail in the second chapter. In the following sections I describe

---

[5] Cfr. J. Copeland, *Artificial Intelligence: A Philosophical Introduction*, Oxford: Blackwell, 1993, p. 9.

[6] Cfr. M. Kantrowitz, "Milestones in the Development of Artificial Intelligence" at ftp://ftp.cs.cmu.edu/user/ai/pubs/fa-qs/ai/timeline.txt, Carnegie Mellon, Pittsburgh, PA, 1994.

several contexts in which the term "Artificial Intelligence" is used, highlighting the differences and similarities of use among the cases.

1.1.2. "Artificial Intelligence" in popular culture

On the popular front, "Artificial Intelligence" is associated with dazzling exhibits with computers that have appeared on the world stage. Perhaps the most spectacular exhibit has been the triumph of Deep Blue over the world chess champion, Gary Kasparov in the first match of a six-game, full-length regulation match sponsored by the Association for Computing Machinery in February of 1996. The encounter was organized in celebration of the fiftieth anniversary of the computer. By his own account, the world chess champion was playing for the whole human species.[7] Skilled chess players must possess the ability to make difficult calculations and recognize a seemingly endless number of patterns. Chess also requires imagination, intuition, ingenuity and endurance.[8] It challenges the cerebral ability of the human person and a player who consistently triumphs over others is generally considered to be intellectually superior. If a machine can defeat a champion at a game that requires such intellectual ability, what does this say about our "unique" human qualities?

Other landmark exhibits of thinking computers appear in the realm of science fiction books and movies. Notable mention must be made of the super computer HAL-9000 in Stanley Kubrick's film: *2001: A Space Odyssey*, launched in 1968. The film was revolutionary and rescued the genre of science fiction from its demise. The script writers Kubrick and Arthur Clarke invested a great deal of scientific research into the movie and sought to be in harmony with what they thought were real possibilities in science. HAL appears even more human than his empty emotionless human companions. When Bowman manages to eliminate HAL, HAL pleads and confesses to be afraid making the spectator doubt whether what he is seeing is a simple disconnection or an assassination.

Aside from these spectacular exhibits the ordinary layman is constantly bombarded with new possibilities in computing and communications. A single "smart" card can be a drivers permit, student identity card, passport, calling card, credit card etc., all rolled into one and valid anywhere in the world. Networking software "intelligently" routes and reroutes packets of information along cables to the destination you specify. We also now have inexpensive personal computers that can support high-performance automatic speech recognition software (ASR). Kurzweil, Dragon Systems and IBM, for example, offer highly publicized commercial products that have proven useful though their

---

[7] Cfr. M. Campbell, "How HAL Plays Chess" in D. G. Stork, ed., *Hal's Legacy: 2001's Computer as Dream and Reality*, p. 78.

[8] Cfr. M. Campbell, "How HAL Plays Chess", pp. 75-100.

limitations are patent. These systems, however, are a long way from recognizing voice with the efficiency, flexibility and intelligence which humans display.[9]

In addition to these advances computers and computer gadgets more and more have a human appeal. Commercial programs for learning, for example, are decisively interactive, permitting dialogue between the user and the machine. The machine responds using images, text and sound in a manner that is particularly adapted to the way humans communicate and learn. Interactive technology in the public forum has given rise to computer games that are ever more realistic, lending additional prestige to the idea of intelligent machines. Added to all this, the media is a welcoming forum for enthusiastic scientists to publicize their futuristic predictions, hopes and desires for AI.

Given this panorama, what significance does the ordinary layman attach to the term "Artificial Intelligence"? The person on the street understands that computers or fancy electronics are involved doing incredible things. His or her understanding probably does not go much deeper than this. He or she gives little relevance to the meaning of the terms and tacitly accepts referring to artifacts as "intelligent" or "smart".

1.1.3. "Artificial Intelligence" in Applied AI

Those who work in applied artificial intelligence use the term "Artificial Intelligence" in relation to more specific concepts or theories related to their concrete project. AI labs generally focus on simulating or replicating in machines a particular type of intelligent behavior. An important area, for example, is that of recognizing speech. In these laboratories, artificial intelligence is related to complex analysis of speech signals in which the relationship between the various frequencies in an utterance can enable a computer to decipher the words pronounced.[10] Other AI labs center around computer vision. Here AI is related to the study of how computers can distinguish the shape, parts and orientation of objects by means of an analysis of the variations in intensity of light that is reflected from different parts of the object. This information can then be used for recognition - for example, in identifying a particular person's face in a collection of face images or in recognizing facial expressions.[11]

An area that occupies a special place in applied AI is knowledge representation in computer systems.[12] Applied research in this area is closely associated with theories and

---

[9] Cfr. R. Schank, "How Could HAL Use Language", p. 175.

[10] Cfr. R. Kurzweil, "When Will HAL Understand What We Are Saying? Computer Speech Recognition and Understanding" in D. G. Stork, ed., *Hal's Legacy: 2001's Computer as Dream and Reality*, pp. 131-170.

[11] Cfr. A. Rosenfeld, "Eyes for Computers: How HAL Could See" in D. G. Stork, ed., *Hal's Legacy: 2001's Computer as Dream and Reality*, pp. 211-236.

[12] Cfr. N. Nilsson, *Principles of Artificial Intelligence*, Palo Alto, CA: Tioga, 1980.

models of what knowledge, knowing and reasoning is about. These theories are mostly derived from logic, cognitive science, psychology and philosophy. In knowledge representation, on the technical end, three basic approaches can be identified: case-based reasoning or CBR, rule-based reasoning, sometimes called expert systems, and connectionist reasoning.[13]

The basic idea in CBR is that the program has stored problems and solutions. Then when a new problem arises, the program tries to find a similar problem in its database by finding analogous aspects between the problems. The problem is that it is very difficult to know which aspects from one problem should match which ones in any candidate problem in the other, especially if some of the features are absent. In rule-based or expert systems, the programmer enters a large number of rules. The problem here is that it is wishful to try to anticipate every possible input. It is extremely problematic to be sure that there are rules that will cover all the circumstances that may arise. Thus these systems often break down when unforeseen problems are presented. They are very 'brittle'.[14] Connectionists use learning rules in big networks of simple components - loosely inspired by nerves in a brain. Aside from language in its diverse aspects (recognition, generation, understanding etc.), computer vision and knowledge representation, AI labs investigate numerous other areas such as robotics, emotions and interfaces.

In the wide variety of AI projects, a common element can be identified in the manner in which the researchers in the respective areas evaluate the progress that they have achieved. How does it stand up to a human like way of doing things? This evaluation, however, is not a vague comparison according to the layman's understanding of what human performance in language or vision for example might be. Each branch of AI research has specific and at times elaborate criteria for evaluating their progress. The criteria, however, generally refer to specific measures that have human performance as their reference. AI labs that do research in Automatic Speech Recognition (ASR), for example, typically refer to three specific criteria: effectiveness in recognizing continuous speech, vocabulary size and understanding, and speaker independence.[15] These criteria have specific and well defined meanings in language recognition but clearly refer to what are considered principle elements in the way humans recognize and interpret language. In AI circles that investigate ASR, therefore, a machine recognizes language in a human way when it can interpret fully continuous

---

[13] Cfr. D. G. Stork, "Scientist on the Set: An Interview with Marvin Minsky" in D. G. Stork, ed., *Hal's Legacy: 2001's Computer as Dream and Reality*, pp. 17-18.

[14] Cfr. D. B. Lenat and E. A. Feigenbaum, "On the Thresholds of Knowledge", *Artificial Intelligence* 47 (1991), p. 196.

[15] Cfr. R. Kurzweil, "When Will HAL Understand What We Are Saying? Computer Speech Recognition and Understanding", pp. 131-170.

speech with high accuracy, relatively unrestricted vocabulary and with no previous exposure to the speaker.

Investigators in computer vision, on the other hand, evaluate their progress according to the variety of objects that the machine is able to recognize. A simple domain would be, for example, computer reading of documents where illumination can be controlled and the surfaces in the scene are stationary and flat. A more complex domain, which is still out of the reach of modern technology, would be, for example, distinguishing moving three dimensional objects with varying reflectivity along their surfaces and which are subject to an illumination that changes arbitrarily.[16]

For researchers in Computer Vision, attaining systems with high performance in the complex domains is the first barrier to building systems that "see" in a human like manner. More advanced systems will be necessary to interpret peoples actions, intentions, emotions and behaviors. As can be seen, researchers in computer vision use criteria for evaluating their progress which refer to specific and elaborate concepts and theories. The criteria are entirely different from those used in ASR which was discussed above, where the concepts and theories are different.

The term "Artificial Intelligence" thus has a broad usage in the context of applied research and its full comprehension in a given ambit of research depends on the hypotheses, theories, models and criteria for evaluation in the ambit of investigation in which it is used. Researchers work with more or less success according to the aspect of human like behavior they investigate. The simpler ones are naturally quickly implemented. The more difficult ones require more time, effort and money. As time passes and we know more about the mind, language and human behavior, and more powerful computing tools become available, some hypotheses, models and tests prove unworkable and are abandoned in favor of others that seem more workable.

1.1.4. Human AI and alien AI

The brief discussion above reveals that not all AI researchers necessarily believe that we can say that machines can realize intelligent behavior in the same sense that we say humans realize intelligent behavior. This is so because the many branches of AI research each have their own models, objectives and methods which are not necessarily in accord with the demands of human intelligence. Many projects may have commercial or military objectives for example, and as a result they admit techniques and models that are clearly at odds with what real intelligence is about. This is the case for example of many expert systems such as Mycin and Dendral which have the purpose of facilitating decision making in particular areas of human activity, disease diagnosis in the case of Mycin and spectral chemistry in the case of Dendral. These projects use restricted

[16] Cfr. A. Rosenfeld, "Eyes for Computers: How HAL Could See", pp. 212-234.

domains in representing knowledge and limit queries to these domains. This is at odds with the demands of real intelligence where flexibility of language must be admitted.[17]

In many AI projects, in fact, the tendency is to side step the issue of real intelligence as irrelevant and distracting in the context of the immediate challenges at hand which are often very demanding and complicated. Jeffery D. Ullman for example, author of a widely used manual on database and knowledge-base systems, speaks of the terms "Knowledge" and "Artificial Intelligence" quite negatively. He affirms:

> "'Knowledge" is a tricky notion to define formally, and it has been made trickier by the fact that today "knowledge" sells well. It appears that attributing "knowledge" to your software product, or saying that it uses "Artificial Intelligence" makes the product more attractive, even though the performance and functionality may be no better than that of a similar product not claimed to possess these qualities.'[18]

Ullman wishes to limit his discussion about knowledge bases to objectives and criteria that are within the realm of computer science. He goes on to observe: 'when examined, it appears that the term "knowledge" is used chiefly as an attribute of programming systems that support some form of declarative language.'[19] Having established this, he explicitly distances himself from philosophical discussions and defines the ambit of his study. He declares: 'We shall therefore sidestep the philosophical questions of what "knowledge" is and use the term to refer to systems with declarative, logic-based languages.'[20] Jeffery Ullman is a prestigious expert in knowledge-base systems. His low esteem for the question of real intelligence in artifacts illustrates that researchers in AI themes do not necessarily subscribe to the project of emulating intelligence in machines.

The terms "human AI" and "alien AI" are sometimes used to distinguish approaches to artificial intelligence according to whether the objectives and techniques admitted are in accord or not with the demands of a thinking machine.[21] Each is the outcome of a fundamentally different way of approaching AI. On the one hand, researchers can

---

[17] Cfr. D. B. Lenat and E. A. Feigenbaum, OTK, p. 196.

[18] J. D. Ullman, *Principles of Database and Knowledge-Base Systems*, Vol. 1, Rockville, MD: Computer Science Press, 1989, p. 23.

[19] A declarative language is a logical language for specifying the characteristics of knowledge items and the relationships between them.

[20] J. D. Ullman, *Principles of Database and Knowledge-Base Systems*, p. 24.

[21] Cfr. J. Copeland, *Artificial Intelligence: A Philosophical Introduction*, pp. 207-208. J. Searle uses the terminology "Strong AI" and "Weak AI" to make the distinction that is discussed in this section (Cfr. J. Searle, "Minds, Brains, and Programs" in J. Haugland, ed., *Mind Design II*, p. 183. The article was written in 1980).

attempt to write programs that simulate or mimic what goes on in the human mind. This is human AI. On the other hand, programmers can consider themselves at liberty to use any techniques whatever, irrespective of whether or not these bear a resemblance to human thought processes. This is alien AI.

Granted this distinction, there is the curious phenomenon, however, that a single project can be considered human AI under certain aspects and alien AI under others. This occurs because the labels "human AI" and "alien AI" are placed by cognitive scientists and philosophers who are interested in any given project only to the extent that it serves to test computational models of human behavior. A given project may not necessarily have this as its goal. As we indicated above, in applied AI the goals can be simply technological. Nevertheless, given that the projects in applied AI normally seek to imitate human ways of doing things, they often involve the implementation of computational models of behavior. In as much as these projects implement computational models of behavior they use techniques that bear resemblance to human thought processes and under this aspect they warrant the title of "human AI". These same projects, however, to achieve their own goals, may implement techniques such as domain restrictions and brute searches which introduce inflexibility. Under these aspects the project counts as alien AI. The labels "human" or "alien", as a result, are placed according to the goals for artificial intelligence in the context of cognitive science and philosophy.

1.1.5. "Artificial Intelligence" in Cognitive Science

In cognitive science the term "Artificial Intelligence" is understood as an approach to investigating behavior using computational models. In their impressive survey of AI, Eugene Charniak and Drew McDermott see this as the principle manner in which AI ought to be understood. They define AI in general as: 'the study of mental faculties through the use of computational models.'[22] It is based on the assumption that intelligence can best be analyzed by trying to reproduce it. In its outset, the AI approach contrasted with an older method of studying cognition, that of experimental psychology.[23] Experimental psychology employs behavioral models and theories. These models seek to explain behavior in terms of physiological and biological functions and in terms of social influences and evolution.

There has been a certain amount of antagonism between the two approaches, with the proponents of each pointing out the strengths of their own methodology and the weaknesses of their opponent's. The approach of psychology and the new AI approach, however, have inevitably interacted with each other. Psychologists have borrowed

---

[22] E. Charniak and D. McDermott, *Introduction to Artificial Intelligence*, p. 6.

[23] Cfr. A. Garnham, *Artificial Intelligence: An Introduction*, London: Routledge, 1988, pp. 11-12.

concepts from the AI approach and researchers committed to the new AI approach have taken interest in psychological findings. The uneasy relationship lasted until the late 1970s when many people on both sides felt the need for a more constructive amalgam of these different approaches to the same problem. A new discipline, *Cognitive Science*, came into being, combining the strengths of psychology, AI and other subjects, in particular linguistics, formal logic and philosophy.[24] Cognitive Science attempts to explain the full gamut of intelligent behavior, and the AI approach is one of its principle methods. In the context of Cognitive Science, therefore, the term "Artificial Intelligence" ought to be primarily understood as the study of mental faculties through the use of computational models.

Discussing the aims of AI, Margaret Boden explains:

> 'Computational psychology uses AI concepts and AI methods in formulating and testing its theories. Mental structures and processes are described in computational terms. Usually, the theories are clarified, and their predictions tested, by running them on a computer program.'[25]

The results obtained from running the programs help to understand mental processes and structures and suggest ways ahead for further research.

1.2. Trends in AI

In applied AI it is difficult to distinguish clear trends. Projects in applied AI serve practical interests. They seek to take advantage of computer power, speed and memory to extend our intellectual and sensory capabilities. Each area of research has its defining characteristics according to the aspect of human behavior that it seeks to enhance.

In the context of Cognitive science or more specifically, computational psychology, this is not the case. AI investigators in computational psychology seek to better understand the human mind and human intelligence. The challenge is to devise and test computational models of the mind that perform better than previous models. AI research in cognitive science centers around programs of research or schools of thought about ways to model mental structures and processes and as a result there are well defined trends. These trends are important aspects of the world in which CYC was born. In this section three principle trends will be described - classical AI, connectionism and situated robotics.

1.2.1. Classical AI

Classical AI is the best known type of AI, and is sometimes called Good Old Fashioned AI or GOFAI. The early initiatives in AI to achieve visible success were

---

[24] Cfr. A. Garnham, *Artificial Intelligence: An Introduction*, p. 1.

[25] M. Boden, "Artificial Intelligence" in *Routledge Encyclopedia of Philosophy* vol. 1, London: Routledge, 1998, pp. 485-492.

mostly projects in classical AI. These were programs such as The Logic Theorist and General Problem Solver (GPS) of Newell and Simon which were forefront initiatives in the 1950's when AI took its first steps. Projects in classical AI use formal programming languages such as LISP and Prolog. The input to the programs, once captured, are represented using symbolic data structures such as lists, semantic networks, arrays and frames. Margaret Boden illustrates some typical characteristics that can be found in classical AI systems with the following toy example of a classical AI program:

If thirsty

then set goal to drink.

If current goal is drink and weather is cold

then set goal to prepare tea.

If current goal is prepare tea and there is no kettle

then set goal to seek kettle.

if current goal is seek kettle and not in kitchen

then go to kitchen and locate kettle.

If kettle is empty

then fill kettle with water.

If kettle is full

then put kettle on hob and heat hob and locate teapot.

(and so on).[26]

This toy example illustrates a trait of many programs in classical AI: every action, and every condition for action, has to be explicitly specified. Actions that undo previous actions such as emptying the kettle that was just filled ought to be evaded. Atypical circumstances that modify the normal course of events have to be anticipated and catered for (turn off the hob). Default steps ought to be specified for cases where the anticipated conditions do not apply (hot weather, not thirsty).

Another trait of many classical AI programs which the example illustrates concerns the handling of goals. There is a well defined hierarchy of goals. The program knows what its current goal is at a given point (seek kettle) as well as the higher level goals (prepare tea and drink). The program must be able to jump up to the higher level goal once the lower sub-goals have been achieved or abandoned. Thus in the example above, once the kettle has been located the program adopts "prepare tea" as its current goal. Supposing the program above controls a simple robot, then there must be procedures for carrying out the tests (is it cold?, is the kettle full?) and for executing actions such as go to, seek and locate.

---

[26] M. Boden, "Artificial Intelligence", p. 487.

The techniques of classical AI permit relatively transparent models whose workings can be well understood by inspecting the program. The explicit approach makes it easy to "debug" the program and to add on new routines in order to improve its capabilities. A further advantage of classical AI is its ability to represent hierarchical structures of goals and to define strong (exceptionless) problem constraints. This facilitates the writing and execution of inference procedures to make the program more responsive. Research in classical AI has led to a proliferation of logical systems and their implementation in computers. Some logical techniques such as means-ends analysis, predicate calculus and relational systems have enjoyed a great deal of success. Many others have proven unworkable and there are at present, many experimental systems in progress.

1.2.2. Connectionism

Advocates of classical AI have been severely criticized for ignoring the physical properties and organic operations of the brain in the formulations they propose as models of thought. Another school, connectionist AI, develops its models based on the findings of neuroscience about the physical operations of the brain.[27]

Connectionist work in AI was inspired by a seminal article in 1943, written by Warren McCulloch, neurophysiologist, and Walter Pitts, mathematician, entitled "A Logical Calculus of the Ideas Immanent in Nervous Activity".[28] The authors showed that simple combinations of idealized neurons could act as symbol manipulators. Each neuron could be regarded as either On or Off with no in-between states. What counts therefore is whether the neuron is firing or not. The intensity of the burst or fluctuations in the cell's chemical structure, for example, are ignored for simplicity. Modeled as a simple On/Off switch, a neuron can be viewed as a device for physically realizing one or other of the two binary symbols 0 or 1. Enlarging on this theme, McCulloch and Pitts explained how small groups of neurons could function as simple manipulators. For instance, a McCulloch-Pitts neuron with two inputs which fires if and only if both inputs were firing constitutes the classical AND operation. A neuron with two inputs which fires if only one of the inputs fire constitutes the classical OR operation. A neuron which fires if some specific input is not firing constitutes the classical NOT operation in binary code. Since every truth function can be expressed with NOT and OR alone, McCulloch and Pitts were able to show that every function of prepositional calculus is realizable by some neural net. Towards the end of their article they compared a net of interconnected neurons and a Turing machine.

---

[27] Cfr. D. E. Rumelhart, "The Architecture of Mind: A Connectionist Approach" in J. Haugland, ed., *Mind Design II*, Cambridge, MA: MIT Press, 1997, pp. 205-232. The article was written in 1989.

[28] Cfr. M. A. Boden, ed., *The Philosophy of Artificial Intelligence*, Oxford: Oxford University Press,1990, ch. 1.

McCulloch and Pitts were writing just before the computer age began in the late forties. When the computer age took off in the late fifties and sixties, however, classical AI occupied center stage with impressive exhibits of intelligent programs such as the Logic Theorist and the General Problem Solver. Researchers who came into AI during the sixties and seventies had little motivation to attempt to construct new types of machines based on the functioning of the human brain. Frank Rosenblatt, one researcher who did take neural simulation seriously, was severely criticized by Marvin Minsky and Seymour Papert in 1969. The authors demonstrated that there were important limitations on what Rosenblatt's simple networks ("perceptrons") could compute.[29] Jack Copeland explains:

> 'In their book *Perceptrons* Minsky and Papert showed that Rosenblatt's devices are incapable of calculating a particular class of functions (the so-called linearly inseparable functions). Many important functions fall into this class, and by and large the AI community treated this result as the death knell for the perceptron.'[30]

Minsky and Papert's observations marked a drop in interest in connectionist research. Little work was done on connectionist systems during the 1970's and early 1980's. In the late eighties Rosenblatt's ideas were revived by researchers in AI who strongly felt that AI models ought to resemble the biochemical processes of the brain. A group of investigators was formed to study and test connectionist models in greater detail. The group calls their approach to AI "Parallel Distributed Processing" or PDP.

The basic building blocks of PDP networks are simple units that have adjustable firing thresholds. A unit in a connectionist network can be an input unit, an output unit or a hidden unit. Input units can receive signals directly from sources external to the network. The output units can directly send signals outside the network. The hidden units mediate the transfer of signals between the input and output units. The connections between units are the direct causal routes along which units send signals to their output units. Units process information by computing an activation function and an output function. By means of local interactions among units, global consequences ensue. The behavior of a network as a whole is a consequence of the pattern of connectivity exhibited by its units at a given time and the global activation state of the network at that time. The latter is indexed by a vector which consists of an ordered list of the activation values of each unit in the network at that time. Network information processing is characterized as the evolution through time of global patterns of activation. The set of all possible global activation states of a network is its activation space. The vectors that index possible global states of the network pick out points in the

---

[29] Cfr. M. Minsky and S. Papert, *Perceptrons: An Introduction to Computational Geometry*, Cambridge, MA: MIT Press, 1969.

[30] Cfr. J. Copeland, *Artificial Intelligence: A Philosophical Introduction*, p. 277.

activation space, and the temporal evolution of the network is characterized as a trajectory through that space.[31]

Representations within a network can be of one or two kinds: local or distributed. A representation is local if it is or is realized by an individual unit in a certain activation state. A representation is distributed if it is or is realized by a pattern of activation over a group of two or more units. The patterns of activation over input units function like questions posed to the network. Unlike classical AI, there is no predetermined format of representation, nor are there addressable storage locations. The patterns of activation over input units are not interpreted according to some declarative language as in classical AI. They are simply patterns that give rise to changes in the pattern of connectivity throughout the entire network.

Neural networks lack explicit inference mechanisms as in classical AI because there is no corresponding declarative language for specifying input and addressable storage locations are not available. The output, like the input, is a pattern of activation over output units. A network computes the input/output function by computing the corresponding vector-to-vector function in the manner described above. Unlike classical mechanisms, networks do not explicitly store the answers as data structures. The answers, rather, are considered to be *implicitly* contained in the pattern of connectivity. Justifying brain-style computation, David Rumelhart explains the importance of this point:

> 'A further consideration differentiates our models from those inspired by the computer metaphor - that is, the constraint that all the knowledge is *in the connections*... Indeed it is the connections - or perhaps the rules for forming them through experience - that primarily differentiate one model from another. This is a profound difference between our approach and other more conventional approaches, for it means that almost all knowledge is *implicit* in the structure of the device that carries out the task, rather that *explicit* in the stated of units themselves. Knowledge is not directly accessible to interpretation by some separate processor, but it is built into the processor itself and directly determines the course of processing. It is acquired through tuning of connections, as they are used in processing, rather than formulated and stored as declarative facts.'[32]

The patterns of activation over output units can be construed as data structures. Such patterns however are produced in response to the network's pattern of connectivity, rather than retrieved from some storage location in the network. A network's memory is in its pattern of connectivity.

---

[31] Cfr. D. E. Rumelhart, "The Architecture of Mind: A Connectionist Approach" in J. Haugland, ed., *Mind Design II*, pp. 210-220.

[32] D. E. Rumelhart, "The Architecture of Mind: A Connectionist Approach", p. 208.

The pattern of connectivity exhibited by the units in a network works like an "implicit" look-up table program. A look-up table program looks up unique answers to questions posed to it. The look-up table, however, need not be pre-programmed into the network. The network can develop it in response to training. It is taught to recognize a pattern, for instance, by being trained on examples of the pattern. Neural networks can learn in this way. During training, the strengths of the connections between units change according to learning rules. The generalized delta rule, typically called "back-propagation", is currently the most widely used learning rule in connectionist research. It is important because it dispelled concerns about the possible computational limitations of network architectures raised by Minsky's and Papert's critique of perceptrons.

Networks are especially proficient at solving what are called connected problems. These are problems such as finding shortest routes which do not divide into independently solvable subproblems. There are also classical methods for solving these problems. As the connected parameters of the problem increase, however, there is an exponential increase in the computational resources required to solve them using classical methods. Connectionist networks are able to solve such problems more efficiently.

Networks are also proficient at pattern-recognition tasks. One network that has received a great deal of attention is Sejnowski and Rosenberg's NETtalk. NETtalk learns to associate letters with phonemes. The network drives a synthesizer that produces pronunciations of the phonemes. After sufficient network training, the synthesizer sounds like a robotic voice literally reading English texts presented to it.[33]

While PDP systems have shown signs of promise, there are many drawbacks. Such systems were inspired by neuronal connections in the brain. Extant connectionist networks, however, are unlike real brains in many ways. The widely used back-propagation algorithm, for example, learns in a fashion that is very unlike real biological processes. Another drawback of first-generation PDP systems is that, while they are able to learn a set of mappings, they do not capture well the function that is implicit in the exemplars they are presented in such a way that they respond properly to cases not yet observed. Unlike classical AI programs, they cannot model hierarchical structure or sequential processing. Many aspects of language and problem solving require both these features. The sophistication of PDP products today are far behind those that have been constructed using classical techniques and it remains unclear the potential role of connectionism in the computational theory of cognition.

---

[33] Cfr. J. Sejnowski and C. Rosenberg, "Parallel Networks That Learn to Pronounce English Text", *Complex Systems* I (1987), pp. 145-168.

1.2.3. Situated robotics

Some see in connectionism an alternative to the classical conception of cognition as rule-governed symbol manipulation. Paul Smolensky, for example, upholds that a connectionist cognitive architecture would contain sentence-like representations as a phenomenon emerging from the network.[34] Others point out that connectionist research is a dead end road if there are no clear means for carrying out the logical processes that characterize thought. In this view, connectionism is best regarded as an interesting way to implement logical languages.[35] There is an intense debate over the issue.

Side-stepping the debate, a group of researchers attack both classicism and connectionism for having overlooked many important aspects of human cognition in their research programs. This partial focus, they diagnose, is the cause of the current impasse. These investigators feel that AI ought to be approached from the ground up, beginning with sensation and having systems that are complete at each stage. Advocates of this view devise and test models by building autonomous mobile robots which interact with their environment. Such investigation is termed "situated robotics" and it is characterized in a special way by its anti-representationalism stance.

Rodney Brooks, a leading figure in situated robotics research, explains that there are four key ideas behind this new style of AI research which he prefers to call more specifically "behavior-based robots". These are: situatedness, embodiment, intelligence and emergence.[36] Situatedness means that the system participates in the external world. There is a blurring between the knowledge of the agent and the world it operates in. The underlying notion, as Brooks expresses it, is that "*The world is its own best model*" and classical representation is carefully avoided.

Another idea is embodiment. The notion here is that only an embodied intelligent agent is capable of dealing with the real world. Furthermore, only through physical interaction with the world can any internal symbolic or other system find empirical grounding. Such grounding enables the system to give "meaning" to the processes going on inside itself. Brooks expresses the key idea in embodiment with the phrase: "*The world grounds the regress of meaning-giving.*"

The third idea is intelligence. Intelligence expresses the notion that intelligent behavior begins long before higher-level processes play a part. Situated robotics proposes to look at simpler animals as a bottom-up model for building intelligence.

---

[34] Cfr. P. Smolensky, "Connectionist Modeling: Neural Computation/Mental Connections" in J. Haugland, ed., *Mind Design II*, pp. 233-250. The article was written in 1989.

[35] Cfr. J. Fodor and Z. Pylyshyn, "Connectionism and Cognitive Architecture: A Critical Analysis", in J. Haugland, ed., *Mind Design II*, pp. 309-350. The article was written 1988.

[36] Cfr. R. Brooks, "Intelligence without Representation" in J. Haugland, ed., *Mind Design II*, p. 416. The article was written in 1991.

Brooks explains that once higher-level processes are set aside as the prime component of a robot's intellect, the dynamics of the interaction of the robot and its environment are primary determinants of the structure of its intelligence. To summarize the key idea from intelligence he employs the phrase: '*Intelligence is determined by the dynamics of interaction with the world*.'

The final idea is emergence. Brooks explains:

> 'It is hard to point at a single component as the seat of intelligence. There is no homunculus. Rather, intelligence emerges from the interaction of the components of the system.'[37]

Brooks sees a radical difference between traditional and behavior-based systems in the way in which intelligence emerges. He emphasizes that in traditional AI, the systems that are build center around modules for high-level intellectual functions such as perception, planning, learning, reasoning and so forth. The intelligent behavior of the system as a whole emerges from the interaction of these sort of advanced components. In behavior-based AI, argues Brooks, the order of events is inverted. The modules that are defined are directly behavior-producing. Typically these might include modules for avoiding obstacles, standing up, controlling gaze and so on. Higher-level functions such as perception, planning, modeling and learning emerge from the interaction of the behavior producing components. Brooks expresses the key idea from emergence with the phrase: "*Intelligence is in the eye of the observer.*"

With these key ideas firmly established, Rodney Brooks explains the immediate goals of his research as follows:

> 'I wish to build completely autonomous mobile agents that co-exist in the world with humans, and are seen by those humans as intelligent beings in their own right. I will call such agents *Creatures*.' [38]

Brooks considers the building of Creatures to be an engineering endeavor and proposes an engineering methodology for building Creatures. At MIT Artificial Intelligence Laboratory he and his colleagues have built a series of robots based on the methodology that they have devised. The first of their creature-robots is named *Allen* and the second *Herbert* which was a much more ambitious project. Allen and Herbert operate in an unconstrained dynamic world (the laboratory and office areas in MIT). They successfully operate with people walking by, with people deliberately trying to confuse them, and with people just standing around watching them. Brooks and his team claim that these robots are *Creatures* in the sense that, on power-up, they exist in the world and interact with it, exhibiting determined sorts of behavior. They believe that their robots operate at a level closer to simple insect-level intelligence than to bacteria-level intelligence. Brooks and his colleagues at MIT are pleased with the performance

---

[37] R. Brooks, "Intelligence without Representation", p. 418.

[38] Cfr. R. Brooks, "Intelligence without Representation", p. 401.

of their systems to date. They realize, however, that there is a long way to travel before robots exhibiting higher-level intelligent behavior are fashioned in AI labs.[39]

1.3. The philosophy of AI

As previously discussed, in the philosophy of AI, investigators are concerned primarily with the question as to whether or not it is legitimate to affirm that an artifact such as a computer can be said to think in the same way that we say humans think. In this context "Artificial intelligence" ought to be understood as referring ultimately to the question of the philosophical possibility of a thinking machine and the limits and scope of AI research.

The philosophy of AI is important because it assesses the theoretical assumptions on which AI research is based. If a thinking machine is philosophically impossible, then to what extent can the human or strong AI approach in cognitive science really explain anything about human behavior? If a deeper philosophical analysis reveals that the simulation of human intelligence can explain only little or nothing of real human intelligence then human or strong AI is poorly aimed and doomed to frustration.

The question "can a machine think?" requires profound analysis as to what constitutes thought, its manifestations and requirements. Thought is generally considered to be related to a wide variety of human behavior such as guessing, making plans, analogizing, setting goals, taking decisions and so on. As a result, the term "Artificial Intelligence" in the philosophy of AI is used in the context of whether or not and in what sense, a computer can be said to do all these things. The philosophy of AI deals with these deeper questions and as a result it provides orientation and evaluation criteria for constructing computational models of behavior.

1.3.1. Criteria for intelligence

Human AI supposes that computers can think. At the base of human AI are hypotheses about the nature of thought and tests for intelligence that are understood to refer, not just to a sort of alien intelligence that we wish to emulate, which is the case in many labs, but to human intelligence understood as such. The founding father of human AI is Alan Turing, a British logician and mathematician. In 1950, Turing published an article entitled "Computing Machinery and Intelligence" in the philosophical journal *Mind*. Turing discussed the question as to whether machines can think, during which he catalogued and refuted nine objections to the claim that machinery can think.[40]

In his seminal article, Turing described a laboratory experiment which he claimed can be used to settle the question as to whether a given computer can think. His

---

[39] Cfr. R. Brooks, "Intelligence without Representation" p. 402-420.

[40] Cfr. A. M. Turing, "Computing Machinery and Intelligence" in J. Haugland, ed., *Mind Design II*, pp. 29-56. The article was written in 1950.

experiment is known as the Turing Test and in some form or other, is the point of reference for advocates of human AI. The test involves two humans and the computer under investigation. The basic idea of the Test is that one of the humans, the interrogator, must try to figure out which of the other two participants is the computer. The only means that the interrogator has of communicating with the other person and the computer is via a screen and a keyboard. The interrogator can ask wide-ranging and penetrating questions, and the computer is permitted to do everything possible to force a wrong identification. The experiment is repeated a number of times with different people in the two human positions. The experiment is a version of what Turing calls the "imitation game". A typical version of the imitation game is the case where the interrogator tries to determine which of the other two players is a man and which is a woman. The man does everything to hide his gender. Turing suggests that if the number of successful identifications in the human/computer version of the game is not significantly greater than the average number of successful identifications in the gender version of the game, then it is to be concluded that the computer can think.[41]

Since its birth around 1950, the Turing test was considered an adequate criterion for machine intelligence until it was challenged by the case that has come to be known as the "Chinese Room". The scenario of the Chinese Room was posed by John Searle in 1980.[42] In the scenario, a person is locked in an enclosed room with one entry through which Chinese symbols are sent into the room every now and then. A single exit is also provided through which the same or different Chinese symbols can be sent out. We suppose that the person locked in the room is fluent in English for example but ignorant of Chinese. The occupant, however possesses a book of instructions, written in English, directing that certain specific Chinese characters should be sent out when certain other Chinese characters are sent into the room. If the person in the room were to act in accordance with those instructions, it might appear to those outside the room that he or she understands Chinese. While those outside might be fooled into thinking that the occupant understands Chinese, the occupant is, in fact, ignorant of Chinese. Searle observes:

> 'Whatever formal principles you put into a computer will not be sufficient for understanding, since a human will be able to follow the formal principles without understanding anything'[43]

Searle adapted his argument to artificial intelligence suggesting that the characters sent into the room might be called "input", the characters sent out "output" and the book of instructions a "program". The fact that the input-program-output system is able to fool observers into believing that it understood Chinese does not mean that it really understands Chinese. Similarly, if an input-program-output system were to display

---

[41] Cfr. A. M. Turing, "Computing Machinery and Intelligence", p. 30.

[42] Cfr. J. Searle, "Minds, Brains, and Programs", pp. 183-187.

[43] J. Searle, "Minds, Brains, and Programs", p. 187.

behavior that was indistinguishable from the behavior displayed by a human being, that would not suffice to affirm that the system really possessed intelligence.

Searle's scenario of the Chinese Room illustrates in a graphical way that there is a profound difference between appearance and reality. While it may appear that the input-program-output system is intelligent, this fact is not sufficient to conclude the system has authentic intelligence. The Chinese Room scenario does not prove, however, that the system is not intelligent. Such a proof would require a syllogistic argument based on some proven and necessary principle associated with intelligence. The argument would then need to show that computers are at odds with such a principle and therefore cannot be considered authentically intelligent. Searle's argument, however, does not propose necessary conditions for intelligence. His argument demonstrates in a clear manner that the Turing test is simply insufficient.

1.3.2. Hypotheses about intelligence

Those who uphold that computers can be said to think in the same way that we say that humans think, are careful to clarify that not any form of automation can be rightly deemed intelligent behavior. A car engine, for example, is an advanced form of automation. No one would propose, however, that it is capable of thought. For there to be thought, external behavior is not enough. Certain conditions have to be met in the internal operations of the artifact. The challenge that follows is to specify those conditions. The question at the heart of AI is precisely this: What conditions have to met in the internal operations of an artifact for there to be intelligent behavior? The solution that is adopted constitutes the foundation for the ensuing research activity. If the artifacts that implement a particular solution perform well, then we can conclude that the underlying theory is well aimed. In AI, three dominant solutions can identified which correspond to the three dominant trends in AI that were described above - classical AI, connectionism and situated robotics.

1. The Physical Symbol Systems Hypothesis

Researchers in classical AI who advocate the possibility of thinking machines have generally subscribed to a fundamental presupposition of how the internal processes of a thinking thing work. This presupposition is the Physical Symbol Systems Hypothesis of reasoning which was advanced by Allen Newell and Herbert Simon in 1976.[44]

The Physical Symbol Systems Hypothesis maintains that the behavior peculiar to thinking things can be explained in terms of symbol manipulation. The hypothesis separates thinking from the biological processes with which it is normally associated. Underlying the hypothesis is an important and well known distinction drawn by C.S.

---

[44] Cfr. A. Newell and H. Simon, "Computer Science as Empirical Inquiry: Symbols and Search" in J. Haugland, ed., *Mind Design II*, pp. 81-110. The article was written in 1976.

Peirce between what are called symbol *types* and symbol *tokens*. The ASCII code for the letter "A" for example, is 1000001. Here the symbol of *type* "A" is represented using seven binary *tokens*. Instead of a string of 1s and zeroes, we can imagine a series of electrical states in the port of a computer that are either on or off. Here the symbol type is the same but the tokens are constituted in a different manner. In the latter case they are constituted by electrical states. This fact is summarized by saying that symbol types are *multiply realizable*: The same symbol type can be physically realized in any number of ways. The symbol of type "A", as a result, continues to be of type "A" whether it is represented by means of electrical states or connections between neurons (as in the human mind). In this light we can justly say that computers manipulate symbols. Computers shuffle around symbols in accordance with the instructions contained in its program. The recipe proposed by mainstream AI for building a machine that can be said to think is summarized by Jack Copeland as follows:

> '1 Use a suitably rich, recursive, compositional code to represent real world objects, events, actions, relationships etc.
>
> 2 Build up an adequate representation of the world and its workings (including human creations such as language and commerce) inside a universal symbol system. This 'knowledge base' will consist of vast, interconnected structures of symbols. It must include a representation of the machine itself and of its purposes and needs. Opinions differ and to whether programmers will have to 'hand craft' this gigantic structure or whether the machine can be programmed to learn much of it for itself.
>
> 3 Use suitable input devices to form symbolic representations of the flux of environmental stimuli impinging on the machine.
>
> 4 Arrange for complex sequences of the universal symbol system's fundamental operations to be applied to the symbol structures produced by the input devices and to the symbol structures stored in the knowledge base. Further symbol structures result. Some of these are designated as output.
>
> 5 This output is a symbolic representation of appropriate behavioral responses (including verbal ones) to the input. A suitable robot body can be used to 'translate' the symbols into real behavior.'[45]

The Symbol Systems Hypothesis is simply this: The recipe is correct. The hypothesis refers to a universal symbol system. This stresses the fact that the symbol types are the same in the world of humans and the computer world. Symbol types are not limited to alphanumeric characters. They can be for example, primary feelings such as joy or sadness and make up higher activities such as understanding and reasoning.

Using the Chinese Room scenario described above, John Searle has seriously challenged the validity of the Physical Symbol Systems Hypothesis. He argues that there is no reason to suppose that symbol manipulation explains understanding because

---

[45] J. Copeland, *Artificial Intelligence: A Philosophical Introduction*, p. 80.

the English speaker in the Chinese room can move around symbols just as well without understanding what the symbols mean. Searle affirms:

> 'As long as the program is defined in terms of computational operations on purely formally specified elements, what the example suggests is that these by themselves have no interesting connection with understanding. They are certainly not sufficient conditions, and not the slightest reason has been given to suppose that they are necessary conditions or even that they make a significant contribution to understanding.'[46]

Searle's point of view is that programmed computers following formal instructions understands what a car or an adding machine understand: exactly nothing. The computer's understanding is zero.

2. Replacement connectionism

Researchers in replacement connectionism propose alternative internal conditions for intelligent behavior. Instead of symbol systems and logical languages, they suggest that intelligence can be explained in terms of changing patterns of connectivity in networks of neurons. Replacement connectionists such as Paul Smolensky maintain that a connectionist cognitive architecture would contain sentence-like representations as a phenomenon emerging from the network. He affirms that such phenomenon will emerge without appealing to explicit representations and logical languages. In this way connectionism separates itself from the suppositions of classical AI that are expressed in the Physical Symbol Systems Hypothesis.

At present there is no consensus as to exactly how sentence like structures can emerge from a network. This has greatly jeopardized the connectionist endeavor. Along these lines, Jerry Fodor and Zenon Pylyshyn have posed the most widely discussed challenge to replacement connectionism. They point out that classicism can explain the systematic relationships among ideas that share intentional modes and similar concepts. For example, that someone able to think that "John loves the girl" would normally be able to think that "the girl loves John". They point out further, that classicism can explain the productivity of thought, that is, the fact the people can have a potential infinitude of thoughts, and that it can explain the inferential coherence of thought processes such as the truth preserving relations between premises and conclusions in a valid argument. They challenge replacement connectionists to show how a connectionist architecture could explain such phenomenon without implementing a classical one.[47]

Connectionism faces another serious challenge from John Searle. Referring to the Chinese room example, Searle illustrates that PDP systems do not account for intentionality. He considers the hypothetical case in which a PDP network successfully

---

[46] J. Searle, "Minds, Brains, and Programs", p. 187.

[47] Cfr. J. Fodor and Z. Pylyshyn, "Connectionism and Cognitive Architecture: A Critical Analysis", pp. 328-350.

translates Chinese symbols. It does this after a rigorous process of learning. A given pattern of input provokes a pattern of connectivity in the network which, by means of an output module generates the correct translation. Searle points out that connectionist systems, while closer to the brain, continue to be collections of formal procedures. Instead of formal symbol manipulation, PDP systems mimic the formal structure of the sequence of neuron firings at the synapses. Searle observes that a person who is given the program instructions about which synapses to turn on and which to turn off and how to go about adjusting the weights for firing, can mimic the entire process of learning and the subsequent process of translation without understanding a word of Chinese. Searle points out that we may be tempted to argue that the person does not realize true parallel processing as he adjusts synapses one at a time and that understanding is in the pattern of connectivity of the entire system. Searle considers this position invalid. He argues that in principle the person can internalize the formal structure of the PDP network, and do all the neuron firing and adjusting in his or her imagination, exactly mimicking the real system. The person may produce correct answers but has no understanding whatsoever of Chinese. Furthermore, we know from the Church-Turing Thesis, a mathematical result, that any computation that can be carried out on a neural net can be carried out on a symbol-manipulating machine. PDP systems carve off the formal properties of the nervous system, but the causal properties that explain understanding are left behind.[48]

Despite the serious difficulties that replacement connectionism faces, there is a great deal of speculation about the consequences for folk psychology should our cognitive architecture prove entirely connectionist. There is nothing approaching a consensus on this point. Some philosophers argue that connectionism could vindicate folk psychology.[49] Other philosophers argue that it would entail the elimination of many folk-psychological notions such as beliefs, memories, desires, hopes, fears, and other propositional attitudes.[50] One obstacle to evaluating such disputes is that it is unclear exactly what folk psychology entails. The most serious obstacle to adjudicating these disputes, however, is that there is no agreement about what a complete connectionist cognitive architecture entails. As previously mentioned, a popular view is that a cognitive architecture ought to contain sentence-like representations as a phenomenon emerging from the network. How this can come about is poorly articulated. There is no consensus about what kinds of network activity would count as realizing mental operations nor what sorts of patterns of connectivity would count as realizing beliefs,

---

[48] Cfr. J. Searle, "Minds, Brains, and Programs", pp. 193-195.

[49] Cfr. T. Horgan and J. Tienson, "Connectionism and the Commitments of Folk Psychology", *Philosophical Perspectives* 9: 127-152, (1995).

[50] Cfr. W. Ramsey, S. Stich and J. Garon, "Connectionism, Eliminativism, and the Future of Folk Psychology" in J. Haugland, ed., *Mind Design II*, pp. 351-376. The article was written in 1990.

intentions, desires and so on. The impact on folk psychology if replacement connectionism is correct is a discussion that is largely conjectural.

3. Anti-representationalist robotics

Researchers in anti-representationalist robotics propose a third alternative regarding the internal conditions for intelligent behavior. Investigators in this field such as Rodney Brooks, emphasize that, unlike classical systems, the modules that are defined are directly behavior-producing. Behavior-based robots might include modules for moving about, avoiding obstacles, moving things out of the way, and so on. Higher-level functions such as perception, planning, modeling and learning emerge from the interaction of the components. As mentioned above, Brooks expresses the key idea from emergence with the phrase: "*Intelligence is in the eye of the observer.*"

John Searle commends anti-representationalist robotics because it tacitly concedes that cognition is not solely a matter of formal symbol manipulation. Researchers in behavior-based robots add causal relations with the outside world. Searle argues, however, that these systems are also incapable of understanding. To illustrate his argument, Searle again refers to the Chinese Room scenario. In this case, confronted with Chinese symbols, the robot gives a behavioral response. Searle argues that if the computer inside the robot is replaced with a person, the person is capable of realizing the behavior exhibited by the robot without understanding anything at all. All he or she does is manipulate formal symbols. Searle writes:

> 'I am receiving "information" from the robot's "perceptual" apparatus, and I am giving out "instructions" to its motor apparatus without knowing either of these facts. I am the robot's homunculus, but unlike the traditional homunculus, I don't know what's going on. I don't understand anything except the rules for symbol manipulation.'[51]

He concludes that intentional states entirely escape the possibilities of behavior-based robots.

1.3.3. The meaning of "thinking things"

Searle's scenario suggests that even if we were to build and program a super computer that passes the Turing Test, we would still not be justified to claim that it is authentically intelligent. It continues to be that case, however, that no such super computer has ever been built. Furthermore, as AI advances, it becomes increasingly clear that to build such a system will be exceedingly complicated and perhaps simply out of the reach of human possibilities. Many new things have been learnt over the long years of AI research since the sixties and one of those things is that it is extremely difficult to build systems that imitate human behavior, and in particular, systems that

---

[51] J. Searle, "Minds, Brains, and Programs", p. 193.

imitate behavior closely linked with higher-level activities such as understanding, reasoning, setting goals, learning, guessing, lying and so on.

The Turing test proposes a very general and, in large part, non scientific test. In order to have criteria more in accord with the models that are found in AI labs today, some authors propose that we ought to review the way we talk about thinking things so that a door may be opened to the scientific and technical possibility of building artifacts that literally think. Advocates of this view suggest that we can resolve the question as to whether or not artifacts can be said to literally think by changing the way we use language to speak about artifacts.

1.3.3.1. Dennett's intentional stance

Daniel Dennett is a prominent figure in the elaboration and promotion of a linguistic approach to resolve the question as to whether artifacts can think. Dennett suggests that the question can be addressed by looking at our use and understanding of *beliefs*. Dennett puts forth the thesis that while belief is a perfectly objective phenomenon in the sense that a belief can be objectively judged to be true or mistaken, it can be discerned only from the point of view of one who adopts a certain *predictive strategy*, and its existence can be confirmed only by an assessment of the success of that strategy. Dennett stresses the fact that people speak and act according to their beliefs and this can be considered a strategy. Some work better than others. A strategy that works well is worth adopting. To resolve the question at the heart of the philosophy of AI, Dennett proposes a strategy for speaking about thinking things that gets around current deadlocks. He calls his strategy the *intentional strategy* or adopting the *intentional stance*.

To clarify his position, Dennett explains that there are different types of strategies according to the problems that we wish to resolve. A *physical stance* to predict the behavior of a system, for example, would be to determine its physical constitution and the physical nature of elements that intervene in its operation. One applies the laws of physics and chemistry to predict the resulting behavior. In other occasions the physical makeup of a systems does not suit one's goals. Such is the case, for example, when a user relies entirely on the operating instructions to drive a car or set an alarm clock. Dennett call this a *design stance*. Dennett proposes that yet another strategy can be adopted which opens the doors of intentionality to artifacts. This is the *intentional stance*. Dennett describes it as follows:

> 'Here is how it works: first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and purpose. Then you figure out what desires it ought to have, based on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its

beliefs and desires and will in many - but not all - instances yield a decision about what the agent *ought* to do; that is what you predict the agent *will* do.'[52]

Dennett firmly upholds that the strategy works well. Throughout his writings Dennett habitually assigns purposes, desires, goals, beliefs and so on to artifacts. Commenting on Deep Blue's victory over Gary Kasparov in the first game of their 1996 championship match, Dennett affirms:

'Deep Blue beat world chess champion Gary Kasparov by discovering and executing, with exquisite timing, a withering attach, the purposes of which were all too evident in retrospect to Kasparov and his handlers. It was Deep Blue's sensitivity to those purposes and a cognitive capacity to recognize and exploit a subtle flaw in Kasparov's game that explain Deep blue's success. Murray Campbell, Feng-hsiung Hsu, and the other designers of Deep Blue didn't beat Kasparov; Deep Blue did.'[53]

In Dennett's view, Deep Blue is an intentional system. Its behavior is predictable and explainable if we attribute to it beliefs and desires and the rationality required to figure out what it ought to do in the light of those beliefs and desires.

While his writings have enjoyed popular appeal, Dennett's strategy leaves aside an important aspect of beliefs. Beliefs are not merely predictive strategies; they can be objectively judged true or mistaken, founded or unfounded, reasonable or fictitious. The Chinese room scenario articulated by John Searle illustrates that it is an error to attribute beliefs and desires to artifacts. It is an unfounded belief to suppose that artifacts that merely perform formal operations really have desires, goals and their own beliefs. Searle illustrates that they know nothing.[54]

1.3.3.2. Copeland's decision-making strategy

How can a thinking thing be objectively defined? If we possessed a clear definition then it would be simple to determine whether an appropriately programmed computer is the sort of thing that is capable of thought. The problem, however, is not the same as defining an object such as a plant or a phenomenon such as rainfall. In these cases we appeal to the experimental sciences for defining characteristics and proceed via observation to determine whether these characteristics are present or not. In the case of a thought, no one has ever directly observed an object or phenomenon called "thought".

---

[52] D. Dennett, "True Believers: The Intentional Strategy and Why it Works" in J. Haugland, ed., *Mind Design II*, pp. 57-80. The article was written in 1981.

[53] Cfr. D. Dennett, "When HAL Kills, Who's to Blame? Computer Ethics" in D. G. Stork, ed., *Hal's Legacy: 2001's Computer as Dream and Reality*, Cambridge, MA: The MIT Press, 1997, pp. 351-365.

[54] Searle directly addresses the Deep Blue case by formulating what he calls the "Chess Room Argument"; Cfr. J. Searle, "I Married a Computer", *The New York Review of Books* vol. 46/6, April 8, 1999.

We perceive behavior which, having certain characteristics, we consider intelligent and thus proceeding from a subject that thinks. Thought in itself, however, escapes the reach of our senses. A plant or rainfall does not. If thought escapes our senses, where can we turn for a definition of a thinking thing which, as in the experimental sciences, can be universally accepted? The philosophy of AI offers a proposal. It is a matter of exercising our power as a linguistic community to *decide* over the purposes for which we use the concept of a "thinking thing". Jack Copeland strongly defends this approach. In his introductory book to the philosophy of AI. He explains:

> 'The question of whether an electronic artifact can think is not at all like the question of whether (for example) an organism that lacks chlorophyll can photosynthesize. The latter is the sort of question that must be settled by observation: having framed the question, we turn to Mother Nature for the answer. The former, however, is a question that can be settled only by a decision on our part.'[55]

Implicit here is the suggestion that the traditional uses and meanings that are associated with the word "thinking" are no longer adequate for the new technological age in which machines display human-like behavior. Traditional uses of the word "thinking" suppose that thought pertains to a faculty of man that transcends matter. Reasoning and reflection may make *use* of internal biological processes, but the fundamental aspects of such behavior suppose a non-material metaphysical principle in man. Copeland suggests that deeper explanations for thought are illusory because they cannot be settled by direct observation. The question as to whether or not an artifact can be said to think needs to be redirected away from metaphysical probing and towards the linguistic community. The question is no longer about what constitutes thought and from there determine whether machines can think. The question now is to decide as to what definition of thought best serves our purposes as a linguistic community in the modern world. Jack Copeland summarizes:

> 'Now for the first time, we are faced with the prospect of artifacts that are more or less indistinguishable from human beings in point of their ability to perform those sorts of activity that *we* perform by, as we say, understanding the issues involved, making plans, solving problems, taking decisions, and so on through a wide range of psychological descriptions. Should we apply or withhold the description *thinking thing* in this new case (together with all the psychological vocabulary that this description brings in its train)? What we as a linguistic community must consider in making this decision is whether the purposes for which we use the concept of a thinking thing are best served by deciding to count an appropriately programmed computer as a thing that thinks, or not. (As Wittgenstein remarks: 'Look at the word "to think" as a tool'.)'[56].

---

[55] J. Copeland, *Artificial Intelligence: A Philosophical Introduction*, p. 52.

[56] J. Copeland, *Artificial Intelligence: A Philosophical Introduction*, p. 54.

Jack Copeland goes on to argue that the purposes for which we use the concept of a thinking thing in the new age of computer technology are best served if we indeed decide to count an appropriately programmed computer as a thing that thinks. Rather than mere tests - a word which suggests a deeper cause - proposals such as that of Turing become respectful scientific hypotheses. They demand the acceptance and respect that is accorded, not simply to a mere test, but to a working definition on which artificial intelligence can progress. Copeland goes on to explain that in order to decide to predicate "thinking thing" of an artifact, the linguistic community must previously or simultaneously decide to predicate of artifacts, all those things associated with thinking, that is, activities such as reasoning, guessing, planning, devising strategies and so on.

The approach that Copeland defends implies a restructure of language that excludes a reference to the peculiar internal activity in human thought which the Chinese Room scenario brings to light. By means of his scenario, Searle illustrates, for example, that intentionality is an essential part of intelligence. Intentionality expresses the reality that symbols have meaning for intelligent subjects. Intentionality refers to an internal process that is beyond digital computers and yet is an essential aspect of thought. Searle affirms:

> 'Formal symbol manipulations by themselves do not have any intentionality: they are meaningless; they are not even symbol manipulations, since the symbols do not symbolize anything. In the linguistic jargon they have only syntax but no semantics. Such intentionality as computers appear to have is solely in the minds of those who program them and those who send in the input and who interpret the output.'[57]

If language is restructured in the way that Copeland suggests, the notion of intentionality as Searle explains will inevitably be lost. This is so because intentionality cannot be predicated of computers. Furthermore, the pragmatic turn in the philosophy of language has confirmed and elaborated the importance of internal activities in intelligent behavior that cannot be reduced to program code.[58] The move suggested by Copeland thus impoverishes language. The notion of intentionality, for example will be lost. Furthermore, many concepts in the philosophy of language today will have to be abandoned and forgotten. His suggestion does not seem an advance and in many ways is simply impractical.

1.4. How AI has fared and the need for knowledge

In the previous sections I outlined the technological, scientific and philosophical context in which CYC was born. Three principle trends were discussed, classical AI,

---

[57] J. Searle, "Minds, Brains, and Programs", p. 199.

[58] Cfr. J. Nubiola, *La renovación pragmatista de la filosofía analítica. Una introducción a la filosofía contemporánea del lenguaje*. Pamplona: Eunsa, 1994, pp. 95-98.

connectionism and situated robotics. CYC is a project in classical AI to devise and test models of intelligence that are knowledge based. CYC addresses the problem of storing common sense which is widely considered the chief challenge facing AI. In this section, I will explain how and why the problem of storing common sense has come to be of primary concern among cognitive scientist and AI researchers. I employ the term "Artificial Intelligence" primarily in the manner in which it is used by Cognitive Scientists, that is, as the study of mental faculties through the use of computational models.

1.4.1. How AI has fared

The 60's and early 70's were golden years for artificial intelligence. Only a few years had passed since the Dartmouth conference and already there were fascinating exhibits of machine intelligence. Computer power and speeds multiplied exponentially making it easier for researchers to experiment with new programming and computing methods and making such technology more widely accessible. Computer labs were set up at major universities such as MIT and Stanford with healthy budgets for investigation. AI projects multiplied. Many of the landmark exhibits in AI pertain to the sixties. It was in 1965, for example that Buchanan, Feigenbaum and Lederberg began the Dendral expert system project. Dendral worked out chemical structures from spectral data. Dendral showed remarkable competence and fueled the AI dream.[59]

The year after Dendral was launched Weizenbaum displayed his psychotherapist program Eliza which is probably the most famous of its type. Eliza administers therapeutic interviews to human beings.[60] Towards the end of the sixties, Pople and Myers began Internist, an aid in the diagnosis of human diseases. Shortliffe's Mycin, a system similar to Internist began around the same time.[61] In the sixties, chess programs began to show superiority over their human rivals and high-school algebra problems are easily solved by systems such as Newell, Shaw and Simon's *General Problem Solver* (GPS).[62] Many of these programs, however, lacked capacities that are normally associated with intelligence. They heavily relied on brute searches to match problems with prefabricated solutions. Intelligence involves knowledge about the surrounding environment, reasoning ability, planning out actions, learning and understanding of goals and motivations.

---

[59] D. B. Lenat and E. A. Feigenbaum, OTK, p. 193.

[60] Cfr. J. Weizenbaum, "ELIZA - a Computer Program for the Study of Natural Language Communication Between Man and Machine", *Communications of the ACM* 9 (1966), pp. 36-45.

[61] MYCIN has enjoyed enormous success. It well documented in: E. H. Shortliffe, *Computer-Based Medical Consultations: MYCIN*, New York: Elsevier, 1976.

[62] Cfr. G. W. Ernst and A. Newell, *GPS: A Case Study in Generality and Problem Solving*, New York: Academic Press, 1969.

Terry Winograd, working at MIT in the sixties and seventies was one of the first to address the problem of how to give a computer these capacities. The result was his program Shrdlu which operates in a simplified world of simulated colored blocks. When told to arrange the blocks in a particular configuration, Shrdlu will devise and execute his own plan of action. His ability to handle natural English instructions is impressive, and he appears capable of discerning the most probable meaning of an ambiguous command. Shrdlu's reasoning powers enable him to answer complicated questions about his table-top world of colored blocks and is capable, to a limited extent, of discussing his own motives.[63]

Another widely praised effort is Shaky, a mobile robot built at the Stanford Research institute. Shaky is equipped with touch sensors and a TV eye. The robot can maneuver among several interconnecting rooms amid a litter of wooden blocks. On being instructed, for instance, to move block X from its position on a platform in room A to a location beneath a window, Shaky would navigate his way into room A from wherever he happened to be, appraise the situation, decide that he needed a ramp to reach the platform, look around, find one, push it up against the platform, climb the ramp, topple the block onto the floor, descend and carry the block to the required position under the window - all without human inventions.[64]

The baffling progress in the sixties and early seventies contributed to a pervading optimism regarding the possibilities of AI. A machine that satisfied the Turing test seemed just around the corner. The evidence for the Symbol Systems Hypothesis seemed convincing and real thinking machines seemed not too far off. In 1965 Newell Simon made the well known prediction that machines will be capable, within twenty years, of doing any work that a man can do. Twenty years have long past since then and the Turing test continues to stand tall above the rubble of what Marvin Minsky calls 'dumb specialists in small domains'[65]. Jack Copeland summarizes:

> 'The achievements of AI research are meager, even toy-like, when matched against the overall goal of a computer that operates at human levels of intelligence in the unruly complexity of the real world. If Shaky is taken out of his artificially simple laboratory environment and put into a real house, he will perform about on the level with a clockwork mouse.'[66]

The enthusiasm of the 60's and 70's has now waned. Roger C. Shank, a key figure in natural language processing since the sixties now regards the aspirations of those

---

[63] Cfr. J. Copeland, *Artificial Intelligence: A Philosophical Introduction*, p. 84.

[64] Cfr. J. Copeland, *Artificial Intelligence: A Philosophical Introduction*, p. 85.

[65] D. G. Stork, "Scientist on the Set: An Interview with Marvin Minsky", p. 27.

[66] J. Copeland, *Artificial Intelligence: A Philosophical Introduction*, p. 86.

years as illusory.[67] Terry Winograd and Fernando Flores severely criticize the early goals of AI as far-fetched and misguided. The authors affirm:

> 'Until the mid-1970s artificial intelligence researchers generally believed they could work simultaneously towards two goals: extending the capabilities of computers, and moving towards an understanding of human intelligence. Work was aimed towards an ultimate synthesis, In the last few years, this view has been questioned. There is a tacit acceptance of the point we have made in this book - that the techniques of current AI are not adequate for an understanding of human thought and language.'[68]

What has happened that such a sober mood has befallen researchers in the AI community? Why has the enthusiasm waned? The AI community is more mature than it was thirty years ago. The exponential progress in the sixties and seventies can be attributed to a large extent to the failure to recognize the deeper aspects of intelligence. As diagnosed by Minsky and Shank in the quotes above, AI activity in the early years centered around problem solving abilities and were restricted to specific, limited domains. Such, for example is the case of chess playing programs which enjoyed roaring success in the sixties an seventies and which contributed to fueling enthusiasm in AI. Simply speaking, AI began with the easy problems.

The later seventies and eighties were a time to take stock of the progress in AI. It was a time to evaluate the evidence for the Symbol Systems Hypothesis and to determine the way ahead to conquer the Turing Test. The programs from the sixties were widely studied, and reviewed. Their internal processes were uncovered and scrutinized, even by skeptics, with a view to seriously consider the claims of machine intelligence. Viewed from the inside, the intelligent programs of the sixties fell far short of real intelligence. The impressive problem solving abilities of GPS turned out to be the result of pre-conceived choices and brute trial and error algorithms.[69] This has little to do with how humans think. Humans make choices based on their understanding of the situation in question and can bring several different strategies to bear. Brute trial and error is often a desperate last resort. GPS lacked world knowledge and the ability to come up with different strategies.

Shrdlu was born to address the world knowledge and representation problem that afflicted GPS. The strategy was to focus research on small, simple, artificial worlds called micro-worlds and to develop techniques for organizing knowledge in them. The hope was that eventually techniques developed in micro-world research would be generalized to cope with the complex real world. The artificial world of Shrdlu is colored blocks. It ought to be said that Shrdlu was truly an impressive programming

---

[67] Cfr. R. Schank, "How Could HAL Use Language", p. 175.

[68] T. Winograd and F. Flores, *Understanding Computers and Cognition*, Norwood, NJ: Ablex, 1986, p. 126.

[69] Cfr. J. Copeland, *Artificial Intelligence: A Philosophical Introduction*, pp. 86-91.

effort. World knowledge and varieties of representations were incorporated. Nevertheless, Shrdlu's promise to be a gateway to dealing with more complex, less defined scenarios as in the real world has proven illusory.

A look inside Shrdlu revealed that his apparent mastery of words concerning his world of colored blocks was grounded on the programming technique of labeling. On being told, for example, that his human friend with which he interacted owned something, Shrdlu simply attached the label OWN: FRIEND to his 'internal representation' of the object in question.[70] When asked if an object was owned, the program simply checked to see if it had such a label. Shrdlu had no understanding of what it meant to own something. Aside from this, the sheer complexity of the program made it extremely difficult to adapt Shrdlu to wider domains. Oversimplification and complexity reduced Shrdlu to a dead end road in AI.

On the connectionist front, PDP systems began to show promise in the early eighties. Network programs performed well in simple recognition and learning tasks. Progress in connectionist AI, however, is stalled due to the lack of a clear means for storing, accessing and manipulating knowledge.

1.4.2. The need for knowledge

The critique of AI projects that took place in the seventies and early eighties made manifest the strengths and weaknesses of the AI projects to that time. Many lessons were learnt and taken to heart. Perhaps one of the most important lessons drawn from the experience of the first two decades of AI is that programs that do not know much, cannot do much. In an article on language understanding in AI, Roger C. Shank explains what he calls the "Illusion of Intelligence".[71] By this he refers to a radical error in AI research since the sixties which assumed that an entity that engages in intelligent actions such as chess playing is therefore intelligent. The mistake, he explains, is to believe that problem solving ability is at the heart of intelligence. Shanks explains that many AI researchers, believing this to be so, have attempted to build programs to solve all kind of problems, including chess. Reducing intelligence to problem solving, however is illusory. Shank explains:

> 'There is a big difference between memorizing a set of rules and learning how to apply them in a right way at the right time and inventing those rules. A human learns to play chess, invents new strategies and gets better with every game. A computer is born a chess player, learns nothing new and always uses the same strategies.'[72]

The peculiar human abilities to learn, create and improve to which Shank refers are beyond systems that specialize in particular domains of what is considered to be

---

[70] J. Copeland, *Artificial Intelligence: A Philosophical Introduction*, p. 93.

[71] R. Schank, "How Could HAL Use Language", p. 175.

[72] R. Schank, "How Could HAL Use Language", p. 175.

intelligent behavior. Such abilities require systems taht incorporate theories and models for general intelligence.

Considering the lessons of the past, many AI researchers consider the question of general intelligence as the central challenge of AI today. Reflecting on the difficulty of language acquisition, Schank affirms:

> 'The essence of the natural language problem is not language at all. A researcher can input definitions into a computer for decades and still never give it the ability to understand human experience... I am simply saying that the problem is harder than it first appears. The problem is not one of language but of knowledge and the acquisition of knowledge. A computer would need to know a great deal to engage in even a simple dialogue.'[73]

Along similar lines, in an article on speech recognition, Raymond Kurzweil expresses the need to build in knowledge at many levels, and highlights that the greatest difficulty is to build in general knowledge of the subject matter being spoken about. He affirms:

> 'Thus lesson number one for constructing a computer system that can understand human speech is to build-in knowledge at many levels: the structure of speech sounds, the way speech is produced by our vocal apparatus, the patterns of speech sounds that comprise dialects and languages, the complex (and not fully understood) rules of word usage, and - the greatest difficulty - general knowledge of the subject matter being spoken about.'[74]

Douglas Lenat observes that of the numerous projects in the early years of AI, those that appear to have enjoyed most success are precisely those projects where a knowledge intensive approach to problem solving was chosen. Such is the case with Shortliffe's Mycin program to diagnose diseases and with Feigenbaum and Lederberg's Dendral expert system project to work out chemical structures from spectral data.[75] Writing in 1989, Lenat remarks:

> 'It is now common to hear that a program for understanding natural language must have extensive knowledge of it's domain of discourse. Or, a vision program must have an understanding of the 'world' it is intended to analyze scenes from. Or even, a machine learning program must start with a significant body of knowledge which it will expand, rather than trying to learn from scratch.'[76]

The need for knowledge has spurred on a great deal of investigation and debate over how humans acquire and use real-world knowledge. This debate has led to radical new

---

[73] R. Schank, "How Could HAL Use Language", p. 183.

[74] R. Kurzweil, "When Will HAL Understand What We Are Saying? Computer Speech Recognition and Understanding", p. 133.

[75] D. B. Lenat and E. A. Feigenbaum, OTK, p. 195.

[76] D. B. Lenat and E. A. Feigenbaum, OTK, p. 196.

ways of approaching AI. Researchers in connectionist systems, for example, stress the role of neuron connections and patterns of activation in the brain in the way knowledge is acquired and used. Investigators in situated robotics underline the importance of participation in the world and of the behaviour producing mechanisms that come into play when a subject interacts with his or her environment. These two trends break away from the classical conception of cognition as rule-governed symbol manipulation in order to better account for more sophisticated pehenomenon in the way humans acquire and use knowledge.[77]

1.5. The problem of storing common sense

From a classical standpoint, the transition from micro-worlds such as Terry Winograd's Shrdlu to reality can be made only if computers can be supplied with portions of the vast store of real-world knowledge that we humans use to get along in daily life. Incorporating knowledge, however, is not simply a question of reading in large volumes of data on diverse subjects with a comprehensive index attached. Knowledge must be arranged in some sort of order inside the machine if it is to be useful. Models and techniques need to be devised that can cater for thousands of different types of objects, many of them abstract such as our notions of time, space, causality, beauty, truth and so on. Furthermore, our interest for these objects does not reduce to a well defined set of attributes. It comprehends thousands if not millions of different ways that they can be used in ordinary language.

The complexity involved in storing common knowledge is illustrated by Susumu Kuno's well known analysis of the simple phrase: "Time flies like an arrow". Susumu Kuno developed one of the first systems to uncover the possible interpretations of short phrases in 1963 at Harvard University. Kuno asked his program what this sentence meant. In what has become a famous response, the computer replied that it was not quite sure. It might mean:

1. That time passes as quickly as an arrow passes.

2. Or maybe it is a command telling us to time the flies the same way that an arrow times flies; that is, time flies like an arrow would.

3. Or it could be a command telling us to time only those flies that are similar to arrows; that is, time flies that are like an arrow.

4. Or perhaps it means that a type of fly known as "time flies" have a fondness for arrows; that is, time-flies like (i.e. appreciate) an arrow.[78]

---

[77] Cfr. sections 1.2.2. and 1.2.3.

[78] Kuno's experiment is described in R. Kurzweil, "When Will HAL Understand What We Are Saying? Computer Speech Recognition and Understanding", p. 135.

This example illustrates that it is impossible to understand the sentence about time (or even to understand that the sentence is indeed talking about time and not flies) without mastery of the knowledge structures that represent what we know about time, flies, arrows, and how these concepts relate to one another. To time, flies and arrows, we must add thousands of other concepts in order to build up a knowledge store of common sense. Many concepts are open to ambiguities such as those that arose in the example. How ought this vast quantity of diverse knowledge to be organized in order to be usable?

Another problem that arises is how to update the knowledge store. In a knowledge storehouse with vast quantities of diverse information, it is extremely difficult to determine how a new piece of information should be entered. What properties of the new items does the system require so that it can later decipher the many ways that they are used in ordinary conversation? As the elements can be as diverse as time, arrows and flies, and their usage ambiguous, it is impossible to reduce the information that is required to a listing on a simple form. To competently interpret "Time flies like an arrow" the system will need to know that flies are not similar to arrows, that like in the sense of *similar to* ordinarily requires number agreement between the two objects that are compared, that there are no such things as *time flies*, that flies have never shown fondness for arrows, that arrows cannot and do not *time* anything, and so on. When the system is updated with the notions of time, flies and arrows, it has to be updated with sufficient information about the nature and use of these terms so that all these realities are taken into account.

The problem of updating is further complicated by the fact that a new piece of information may require changes in the way that the entire knowledge store is organized calling for countless updates in large sections of the storehouse. Such is the case, for example, when long term expectations, goals or plans are changed. Having to migrate to a new country, for example, may imply learning a new language, acquiring new interests and tastes, making new friends, changing our attitude towards ourselves and others, all this together with the innumerable details of preparation that such a move entails. Incorporating a new simple piece of knowledge, 'migrate to a new country', calls for countless updates everywhere else in the complex network of the knowledge storehouse.

On introducing a new item it is necessary to specify either explicitly or implicitly which aspects of one's world changes and which aspects remains the same when the knowledge is incorporated. Effective methods are needed to enable the computer to determine all the updates that must be made, allowing changes to efficiently flow through the vast store of interconnected information. This challenge is often called the *frame problem*. This problem takes its name from an analogy between successive states of the world and frames in an animated cartoon. To describe any frame completely requires a large number of statements in order to explain the elements of the scene and the relationships between them. From one frame to another, however, most of the statements remain the same. Only a few change. The frame problem is to specify which

ones change as the result of an action being performed (i.e. the incorporation of new knowledge).[79]

A third challenge in building systems that incorporate vast quantities of knowledge lies in extracting the knowledge that is needed in a given moment in order to resolve a problem that arises. This is the problem of *relevance*. In order to do this, the system needs to be provided with knowledge about how the knowledge it has can be brought to bear to resolve different types of problems. This type of knowledge is often referred to as meta-knowledge.[80] If an immediate solution cannot be found, the system ought to be able formulate pertinent questions which can help to find solutions. Naturally this is extremely complicated. Cognitive scientist, however, are hard at work finding ways of representing meta-knowledge that may later serve to better understand how we humans realize this sort of activity.

The three problems highlighted above are the central questions surrounding what has become known as the "problem of storing common sense".[81] How is the vast collection of data to be organized inside the machine? How is it to be updated? How can the machine determine which facts might be relevant to solve different types of problems? Experts in cognitive science and artificial intelligence have invested a great deal of time and effort to better understand these problems and to propose solutions. Among the numerous initiatives, the boldest is widely considered to be the CYC project to store common sense, led by Douglas Lenat. Speaking on the need for AI researchers to focus on the deeper problems of knowledge representation that is associated with general intelligence, Marvin Minsky points to Lenat's project as a flagship initiative in AI. Speaking about how AI has progressed, he observes:

> 'Bit by bit, people recognized the severity of the knowledge representation problem, but only *Doug Lenat* took it seriously enough to base a research program on it... Doug renounced trying to make intelligence in a particular domain and I think this is a huge advance. I think Lenat is headed in the right direction.'[82]

Philosopher Jack Copeland considers CYC 'the severest test yet of the Symbol Systems Hypothesis' and views CYC as 'one of the few AI projects to take the problem

---

[79] For a more technical discussion of the frame problem see: S. Hanks and D. McDermott, "Default Reasoning, Nonmonotonic Logic, and the Frame Problem", *Proceedings of the Fifth National Conference on Artificial Intelligence* (1986), pp. 328-333.

[80] Cfr. F. Heyes-Roth, D. A. Waterman and D. Lenat, eds., *Building Expert Systems,* Reading, MA: Addison-Wesley, 1983, pp. 219-239.

[81] Cfr. J. Copeland, *Artificial Intelligence: A Philosophical Introduction*, p. 99.

[82] D. G. Stork, "Scientist on the Set: An Interview with Marvin Minsky", p. 16.

of common sense knowledge seriously'[83]. The following chapter takes a close look at this inspiring project.

[83] J. Copeland, *Artificial Intelligence: A Philosophical Introduction*, p. 102.

Chapter II

The CYC experiment

It was observed in the previous chapter that the general feeling in the AI community is that little attention has been paid to the deeper questions of general intelligence in the AI programs that have been designed and implemented since the sixties. This is due to the fact that the experts in the early days of AI thought that the best way to work towards programs with general intelligence was to begin by writing programs that exhibit what is considered to be intelligent behavior, in small restricted domains such as playing chess and solving mathematical problems. The programs that were designed relied heavily on elaborate problem solving techniques and little on knowledge. The internal operations of these programs employed innovative "tricks" to get around their lack of knowledge; tricks which, though effective, bear little resemblance to what thinking is about.

From the point of view of cognitive science it can be said that the experiments realized in AI to date have invalidated theories and models of intelligence that rely heavily on formal methods and little on knowledge. Considered in this context, CYC is a new and bold experiment to investigate the validity of knowledge centered theories and models of intelligence. If CYC proves competent in many of the activities associated with intelligence (understanding language, learning, making plans, setting goals etc.) and its internal processes are not opposed to what thinking is about, then the internal processes and models on which CYC is built would be a step ahead towards understanding intelligence.

The use of CYC to investigate intelligence, is founded on the Symbol Systems Hypothesis which claims that it is correct to explain behavior in terms of symbol manipulation. If CYC performs well, demonstrating intelligent behavior that is in many ways indistinguishable from how humans behave, then the CYC experiment will serve as strong evidence in favor of the Symbol Systems Hypothesis. If behavior can be explained in terms of symbol manipulation, then it must be admitted that machines can literally think. If CYC performs well, then it may serve to defend the philosophical viewpoint that machines can be said to think in the same way that humans think.

In this chapter I will describe CYC from the perspective of cognitive science, that is, as a scientific experiment. Douglas Lenat himself has often explained CYC from this point of view, most notably in his article "On the Thresholds of Knowledge" which he authored together with Edward A. Feigenbaum. In this article, Lenat and Feigenbaum

describe AI research as scientific inquiry into the nature intelligence[84]. They refer to themselves as scientists and their description of the methodology of AI matches the methodology of the empirical sciences in general. In the first section of this chapter Douglas Lenat, leader of the CYC project, is presented. In the second section the methodology of AI which Lenat and Feigenbaum defend is explained. In the third section the particular hypotheses about intelligence that CYC tests are discussed. In the fourth section, CYC's ontological scheme is described and in the final section a brief history of the project is presented.

2.1. Douglas Lenat, the CYC project leader

Douglas William Lenat was born in Philadelphia in 1950. He grew up there though part of his childhood years were spent in Wilmington, Delaware.[85] His family owned a soda bottling business. From an early age Lenat showed interest in scientific topics. His father died suddenly when Lenat was twelve and a half. Lenat's talents began to show in 1967, when he was a finalist in an International Science Fair. He entered the University of Pennsylvania in 1968. The Vietnam War was at its height and the uncertainty of those times motivated Lenat to accelerate his academic training while his student deferment lasted.

When he started college, Lenat was interested in physics and mathematics. He changed his mind, however, when he was introduced to Artificial Intelligence in a course during his third year. Research in AI was still at an early stage at the time. Lenat was captivated, however, by the challenge of building intelligent machines. After graduating from the University of Pennsylvania, Lenat pursued graduate studies in Computer Science at Stanford University. His 1976 Stanford thesis earned him the bi-annual International Joint Conference on Artificial Intelligence (IJCAI) Computers and Thought Award in 1977.[86]

Throughout the late seventies and early eighties, Lenat continued to forge ahead in AI research as professor of Computer Science at Carnegie-Mellon University and Stanford University. The AM and Eurisko programs which he developed during this period were rich in rules for coming up with plausible new concepts in a variety of domains such as set theory, number theory, naval war games tactics, and programming. His experiments with these systems showed how scientific discovery might be

---

[84] Cfr. D. B. Lenat and E. A. Feigenbaum, "On the Thresholds of Knowledge", *Artificial Intelligence* 47 (1991), p. 204.

[85] Cfr. D. Shasha and C. Lazere, *Out Of Their Minds: The Lives and Discoveries of 15 Great Computer Scientists*, New York: Copernicus, 1995.

[86] Cfr. "The People of Cycorp" at http://www.cyc.com/staff.html, Cycorp, Inc., Austin, TX, 1999.

explained as rule-guided, knowledge-guided search.[87] Though the AM and Eurisko programs ran in several domains, their ultimate limitation was their incompetence in a wider selection of domains. Lenat explains :

> 'The ultimate limitation was not what we expected (CPU time), or hoped for (the need to learn new representations of knowledge), but rather something at once surprising and daunting: the need to have a large fraction of consensus reality already in the machine'.[88]

In 1984, while a professor at Stanford, Lenat proposed a plan for CYC. It would be a very large KB incorporating the facts, rules of thumb, and heuristics for reasoning about the objects and events of modern everyday life. By his estimate, the project would take 20 years to build and cost more than $25 million. Such a project was too expensive for a university to fund, so Lenat sought funding from the newly formed Microelectronics and Computer Consortium (MCC) in Austin, Texas. At MCC, a research consortium for advanced technologies, the development of CYC was supported by several organizations, including the United States Department of Defense, Apple, Bellcore, DEC, DoD, Interval, Kodak, and Microsoft.[89]

During the fall of 1984, Lenat moved over to MCC to assemble a team to begin work on CYC. The core team, now around 45, included philosophers, anthropologists, linguists, scientists, engineers and computer scientists.[90] The goal was to give CYC enough knowledge so that by the late 1990s it could begin to learn by means of natural language conversations and reading.[91] The original ten year funding period for the CYC project at MCC was supposed to end in 1994, but was extended for one more year. In 1995 Lenat founded Cycorp to develop and market commercial versions of CYC. The company has its headquarters in Austin, Texas and continues to forge ahead.[92]

---

[87] Cfr. D. B. Lenat and J. S. Brown, 1984, "Why AM and Eurisko Appear to Work", *Artificial Intelligence* 23 (1983), pp. 269-294. AM and Eurisko were research programs in automatic program synthesis and machine learning.

[88] Cfr. D. B. Lenat and E. A. Feigenbaum, OTK, p. 206.

[89] Cfr. L. Grossman, "Get Smart: How Intelligent Can Artificial Intelligence Be? Lev Grossman Talks To Doug Lenat To Find Out", *Time-Digital Magazine* (October 1998).

[90] Cfr. "The People of Cycorp" at http://www.cyc.com/staff.html, 1999.

[91] Cfr. D. B. Lenat, "From *2001* to 2001: Common Sense and the Mind of HAL" in D. G. Stork, ed., 1997, *Hal's Legacy: 2001's Computer as Dream and Reality*, Cambridge, MA: MIT Press, p. 203.

[92] Cfr. "Cycorp. Creators of the CYC Knowledge Base" at http://www.cyc.com, Cycorp, Inc., Austin, TX, 1999.

2.2. CYC methodology

Lenat clearly spells out the goals and methodology of AI to which he subscribes in the article "On the Thresholds of Knowledge" which was published in the journal *Artificial Intelligence* in 1991. Lenat and Feigenbaum summarize what they consider the major findings and hypotheses of AI to date. They begin by articulating their findings in brief succinct expressions. They go on the explain the basis for their convictions drawing on the experience of over three decades of work in AI. Lenat and Feigenbaum articulate three major findings and hypotheses. The third addresses the question of AI methodology. Lenat and Feigenbaum clearly consider the methodology of AI a central issue. They summarize their central methodological tenets in what they term the "Empirical Inquiry Hypothesis (EH)". The authors state it as follows:

> 'Intelligence is still so poorly understood that Nature still holds most of the important surprises in store for us. So the most profitable way to investigate AI is to embody our hypotheses in programs, and gather data by running the programs. The surprises usually suggest revisions that start the cycle over again. Progress depends on these experiments being able to *falsify* our hypotheses. Falsification is the most common and yet most crucial of surprises. In particular, these programs must be capable of behavior not expected by the experimenter.'[93]

Lenat's itinerary in AI research leading up to CYC exemplifies the EH paradigm of scientific investigation. He was led to CYC as a result of a repeated process of embodying hypotheses in programs, testing them by running the programs and improving the hypotheses based on the results. Particularly relevant was his own research prior to CYC on the AM and Eurisko programs. Lenat recalls:

> 'Though the programs ran in several domains, their ultimate limitation was their incompetence in a wider selection of domains. The ultimate limitation was not what we expected (CPU time), or hoped for (the need to learn new representations of knowledge), but rather something at once surprising and daunting: the need to have a large fraction of consensus reality already in the machine'.[94]

This experience led Lenat to CYC. EH represents the paradigm in which Lenat understands his work and the entire AI effort.

The empirical method which Lenat and Feigenbaum propose has a notable peculiarity: the theories are embodied in programs and data is collected by running programs. This method is peculiar because, unlike other empirical sciences, the experiments are constructed without directly intervening in nature. The experiments are constructed on an artificial platform - computers. To what extent is such a means of experimentation capable of falsifying the hypotheses? In order to evaluate the

---

[93] D. B. Lenat and E. A. Feigenbaum, OTK, p. 204.

[94] D. B. Lenat and E. A. Feigenbaum, OTK, p. 206.

hypotheses, what value can we attach to data produced by programs? These issues are addressed in chapter four where the strengths and weaknesses of CYC are discussed.

2.3. CYC theory of intelligence

CYC tests concrete theories and models of intelligence. Lenat and Feigenbaum explain that the empirical evidence accumulated over the years of AI research has suggested the need for a model of intelligence that is knowledge based. They propose the following definition of intelligence:

> '*Intelligence* is the power to rapidly find an adequate solution in what appears a priori (to observers) to be an immense search space.'[95]

The search space that the authors refer to is the number of items that have to be looked at to discover a solution. To find a word in a dictionary, for example, the search space is the entire dictionary for an observer who does not know how it is structured. For an observer who knows how a dictionary is structured, the search space can be reduced to 1.[96] The authors explain that the power to rapidly find solutions is in function of the knowledge that can be brought to bear to solve the problem. They regard this as the fundamental lesson of AI research to date and summarize it in what they call the "Knowledge Principle".

2.3.1. The Knowledge Principle

Lenat and Feigenbaum assert the Knowledge Principle as follows:

> '**Knowledge Principle (KP)**. A system exhibits intelligent understanding and action at a high level of competence primarily because of the *knowledge* that it can bring to bear: the concepts, facts, representations, methods, models, metaphors, and heuristics about its domain of endeavor.'[97]

Explaining the Knowledge Principle, the authors illustrate that there is a tradeoff between knowledge and search. On one hand, search is time consuming if the space is large. On the other hand, as the search space diminishes, it becomes more difficult to apply the knowledge that is available to find smaller spaces. In such a situation, an intelligent system approaches the limits of its competence.

---

[95] D. B. Lenat and E. A. Feigenbaum, OTK, p. 186.

[96] This example is based on an example that Lenat and Feigenbaum used having to do with the interpretation of spectral data using the expert system DENDRAL. The authors illustrate that knowledge about spectral chemistry reduces the search space to 1 in many cases (D. B. Lenat and E. A. Feigenbaum, OTK, p. 194). The example of the dictionary illustrates the same basic point, while being simpler and more accessible.

[97] D. B. Lenat and E. A. Feigenbaum, OTK, p. 186.

Lenat and Feigenbaum illustrate how the tradeoff between knowledge and search works in modern expert systems.[98] They identify three stages as knowledge is added to an expert system. The first stage represents the "bare minimum" of knowledge that the system needs in order to state the problems that it addresses in a well formed fashion. The two AI experts explain: 'Before you can apply search or knowledge to solve some problem, you need to already know enough to at least state the problem in a well-formed fashion.'[99] In this stage the payoff for adding items of knowledge is very large. A little extra knowledge greatly reduces the search space.

As knowledge is added, a stage is reached in which performance reaches what Lenat and Feigenbaum refer to as the "practitioner level". It is the level of a typical practitioner performing the task. The authors explain that up to this point the knowledge/search tradeoff is strongly tipped in favor of knowledge. The knowledge that correspond to this stage concerns, for example, knowledge of which distinctions to make and which ones to ignore. Beyond the practitioner level is the "expert level". Lenat and Feigenbaum describe this level as follows:

> 'Here, each piece of additional knowledge is only infrequently useful. Such knowledge deals with rare but not unheard-of cases. In this realm the knowledge/search tradeoff is fairly evenly balanced. Sometimes it is worth knowing all those obscure cases, sometimes it is more cost-effective to have general models and "run" them.'[100]

Lenat and Feigenbaum explain, that though it may appear obvious, the tradeoff between knowledge and search was not always appreciated. Referring to the early efforts in AI they observe:

> 'At the time, the pervading view in AI ascribed power to the reasoning processes, to the inference engine and not to the knowledge base. (E.g., consider LT and GPS and the flurry of work on resolution theorem provers.) The knowledge and power hypothesis, supported by Feigenbaum (Dendral), McCarthy (Advise Taker), and a few others, stood as a contra-hypothesis. It stood awaiting further empirical testing to either confirm it or falsify it.'[101]

In their discussion of this point, Lenat and Feigenbaum highlight the enormous success that expert systems have enjoyed compared to other systems. Expert systems are prototypes of the Knowledge Principle in that their design follows the knowledge/search tradeoff paradigm. The authors illustrate how this is articulated in the case of Dendral, an expert system to interpret spectral data.[102] The proliferation of expert systems in engineering, manufacturing, geology, molecular biology, financial

---

[98] Cfr. D. B. Lenat and E. A. Feigenbaum, OTK, p. 190.

[99] D. B. Lenat and E. A. Feigenbaum, OTK, p. 190.

[100] D. B. Lenat and E. A. Feigenbaum, OTK, p. 190.

[101] D. B. Lenat and E. A. Feigenbaum, OTK, p. 195.

[102] Cfr. D. B. Lenat and E. A. Feigenbaum, OTK, p. 194.

services, machinery diagnosis and repair, signal processing, and in many other fields is strong evidence for the Knowledge Principle.[103]

2.3.2. The Explicit Knowledge Principle

If AI wishes to progress, the Knowledge Principle becomes a mandate which the authors express as follows:

> 'The Knowledge Principle is a mandate for humanity to concretize the knowledge used in solving hard problems in various fields. This might lead to faster training based on explicit knowledge rather than apprenticeships. It has already led to thousands of profitable expert systems.'[104]

For knowledge to be useful, it needs to be explicitly defined. The essential characteristics of the knowledge items ought to be specified in a consistent way and it ought to be clearly indicated how the knowledge items relate to each other. When this is observed, the knowledge structure can be separated from the programs that are required. To introduce a new piece of information, only the knowledge structure needs to be modified and this is easily realized. Lenat and Feigenbaum criticize conventional programs that incorporate knowledge in the program code. In these cases knowledge is said to be present in a compiled form. To add a new piece of information the program code must be modified, making changes wherever the new piece of information may have an impact. When the programs are large and complicated this is an extremely costly procedure. For knowledge to be useful it must be declared explicitly. The authors express this conviction in what they call the "Explicit Knowledge Principle". They summarize the argument as follows:

> 'Conventional programs have compiled away much of the knowledge, in order to gain efficiency. The price they pay for this, though, is the high cost of integration a new piece of knowledge into their program once it exists. To put this the other way, you can never be sure in advance, how the knowledge already in the system is going to be used, or added to, in the future. Therefore, much of the knowledge in an intelligent system needs to be represented explicitly, declaratively, although compiled forms of it may be present. We might call this the "Explicit Knowledge Principle".'[105]

How the Explicit Knowledge Principle is applied in CYC is described below in section four where the model is discussed in greater detail.

---

[103] Cfr. D. B. Lenat and E. A. Feigenbaum, OTK, p. 195.

[104] D. B. Lenat and E. A. Feigenbaum, OTK, p. 188.

[105] D. B. Lenat and E. A. Feigenbaum, OTK, p. 191.

2.3.3. Overcoming brittleness

Lenat and Feigenbaum praise expert systems for the valuable evidence that they offer in favor the Knowledge Principle and the Explicit Knowledge Principle. They affirm:

> 'The biggest hurdle of all has already been put well behind us: the enormous local maximum of building and using *explicit-knowledge-free* systems. On the far side of that hill we found a much larger payoff, namely expert systems.'[106]

The Knowledge Principle and the Explicit Knowledge Principle capture the underlying source of the power that expert systems demonstrate. Having weighed the evidence and learnt this lesson, AI is now poised to move beyond the threshold of current expert systems. This ought to be realized building on the principles learnt from the past. Describing the motivation for their paper, the authors affirm:

> 'We see expert technology, too, as just a local maximum. AI is finally beginning to move on beyond that threshold. This paper presents what its authors glimpse on the far side of the expert system local-maximum hill: the promise of a large broad KB serving as the nucleus of crystallization for programs which respond sensibly to novel situations because they can reason more by analogy that by perfect matching, and ultimately, because, like us, they understand the meanings of their terms.'[107]

For the two AI experts, expert systems are only a foretaste of the power that knowledge can bring to AI systems. By exploiting knowledge to its full, AI can move on to new thresholds on the far end of the expert systems hill. The authors take a glimpse at the far end of this hill and suggest that the gains are enormous.

Lenat and Feigenbaum consider that the chief shortcoming of expert systems at present is their "brittleness"[108]. Douglas Lenat offers several examples of this in the opening discussion of his book *Building Large Knowledge Based Systems*.[109] He describes one of his favorite scenarios in which a medical program is told about a rusty old car. It blithely diagnoses measles.[110] Lenat and Feigenbaum diagnose the chief limitation of current expert systems to be their failure to respond sensibly to novel

---

[106] D. B. Lenat and E. A. Feigenbaum, OTK, p. 189. The term "local maximum" refers to techniques that have been favored.

[107] D. B. Lenat and E. A. Feigenbaum, OTK, p. 189.

[108] Cfr. D. B. Lenat and E. A. Feigenbaum, OTK, p. 196.

[109] Cfr. D. B. Lenat and R. V. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project*, Reading, MA: Addison-Wesley, 1990. The book describes the evolution of the CYC project from its inception (Fall 1984) to 1990.

[110] Cfr. D. B. Lenat and R. V. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project*, p. 2.

situations. The primary hurdle facing AI is overcoming brittleness. Describing brittleness in more detail, the authors explain:

> 'A limitation of past and current expert systems is their brittleness. They operate on a high plateau of knowledge and competence until they reach the extremity of their knowledge; then they fall off precipitously to levels of ultimate incompetence. People suffer the same difficulty too, but their plateau is much broader and their slope is more gentle. Part of what cushions the fall are layer upon layer of weaker, more general models that underlie their specific knowledge.'[111]

The authors suggest that expert systems are brittle because the knowledge they possess is specific and limited. Faced with a novel situation they lack layers of knowledge resources that are more general and broad to which they can have recourse. The knowledge incorporated into current expert systems is largely domain-specific. It represents the distillation of experience in a field, 'nuggets of compiled hindsight' as the authors describe it.[112] In their field they can powerfully guide search, but faced with a novel situation they are hopelessly incompetent.

2.3.4. Coping with novelty: The Breadth Hypothesis

Lenat and Feigenbaum take a close look at how we humans deal with novel situations to determine how AI must proceed to overcome brittleness. The authors observe: 'But when confronted by a *novel* situation, human beings turn to reasoning strategies like generalizing and analogizing in real time and (even better) *already having* more general rules to fall back on.'[113] To move beyond the threshold of expert systems, AI must experiment with systems that incorporate these reasoning strategies along with more general knowledge. Lenat and Feigenbaum embody these convictions in what they term the "Breadth Hypothesis". They state the Breadth Hypothesis as follows:

> 'Intelligent performance often requires the problem solver to fall back on increasingly general knowledge, and/or to analogize to specific knowledge from far-flung domains.'[114]

Brittleness falsifies any claim that considers the techniques of current expert systems as sufficient to explain intelligence. A new hypothesis is needed which incorporates the strengths of expert systems while resolving brittleness. Such an hypothesis is what Lenat and Feigenbaum propose in the Breadth Hypothesis. Hypotheses are confirmed or refuted through experimentation. The authors propose that AI can move on to new thresholds by testing the Breadth Hypothesis. The CYC project is the first experiment that tests this hypothesis.

---

[111] D. B. Lenat and E. A. Feigenbaum, OTK, p. 196.

[112] D. B. Lenat and E. A. Feigenbaum, OTK, p. 196.

[113] D. B. Lenat and E. A. Feigenbaum, OTK, p. 196.

[114] D. B. Lenat and E. A. Feigenbaum, OTK, p. 186.

The Breadth Hypothesis is a mandate to build systems that incorporate two fundamental strategies for dealing with novelty. On one hand, the system must have the capacity to fall back on increasingly general knowledge. General knowledge refers to the vast storehouse of knowledge that we take for granted when we engage in conversation or read an article for example. The two AI veterans explain:

> 'Each of us has a vast storehouse of general knowledge, though we rarely talk about any of it explicitly to one another; we just assume that other people already know these things. It they are included in a conversation, or an article, they confuse more than they clarify. Some examples are:
>
> • water flows downhill,
>
> • living things get diseases,
>
> • doing work requires energy,
>
> • people live for a single, contiguous, finite interval of time,
>
> • most cars today are riding on four tires,
>
> • each tire a car is riding on is mounted on a wheel,
>
> • if you fall asleep while driving, your car will start to head out of your lane pretty soon,
>
> • if something big is between you and the thing you want, you will probably have to go around it.'[115]

Lacking these simple commonsense concepts, expert systems are brittle. They make mistakes that often appear ridiculous in human terms.[116] Lenat and Feigenbaum call the storehouse of concepts such as these '*consensus reality* knowledge'[117] How consensus reality is represented in CYC is discussed below in section four.

The other strategy at the heart of the Breadth Hypothesis is reasoning by analogy. Lenat and Feigenbaum sustain that the principle reasoning method that humans employ to deal with novel situations is the analogical method. The two AI veterans recognize that analogy is often used as little more than a literary device that has dramatic power. Analogy however, is much more than a mere literary device. Analogizing to general

---

[115] D. B. Lenat and E. A. Feigenbaum, OTK, p. 197.

[116] The authors defend their position with several examples of such mistakes. They state: 'For instance, when a car loan authorization program approves a loan to a teenager who put down that he had worked at the same job for twenty years; or when a skin disease diagnosis program concludes that my rusted out decade-old Chevy has measles; or when a medical system prescribes an absurd dosage of a drug for a maternity patient whose weight (105) and age (35) were accidentally swapped during the case's type-in.' (D. B. Lenat and E. A. Feigenbaum, OTK, p. 197).

[117] D. B. Lenat and E. A. Feigenbaum, OTK, p. 197.

more distant knowledge provides real heuristic power for dealing with novel situations. The authors advise:

> 'Do not make the mistake we did, of thinking of this reasoning method as little more that a literary device, used for achieving some sort of emotional impact. It can be used to help discover solutions to problems, and to flesh out new concepts; and it can be argued that analogy pervades human communication and perhaps almost all of human thought!'[118]

The authors suggest that, in most part, we reason by analogy to find solutions and to flesh out new concepts. We turn to our storehouse of general knowledge for cases that are similar to the problem that we face, partial-matching from the current situation to more familiar scenarios.[119] Reasoning by analogy, these similar cases can help us to discover new concepts for dealing with the problem at hand. Discussing the analogical method in a book on CYC, Lenat and Guha illustrate how analogy can be used to discover new concepts with the following example:

> 'See for yourself. Here is a partial analogy between treating a disease and waging a war:

Treating a Bacterial Infection	**Fighting a War**
EnemyType: Disease	*EnemyType*: MilitaryForce
EnemyLocal: Bacteria	*EnemyLocal*: EnemyTroops
ProtagonistType: Physician	*ProtagonistType*: Soldier
EnemyProcess: Infecting	*EnemyProcess*: Invading
ProtagProcess: ClinTreating	*ProtagProcess*: MilRepulsing
Preprocess: Diagnosing	*Preprocess*: Spying
Tactics: Vaccination	*Tactics*: MilContainment
Locale: BodyPart	*Locale*: GeographicRegion
EmotionalCharge: Low	*EmotionalCharge*: High'[120]

Lenat and Guha point out that some of the concepts on one side are analogs of those on the other side. In the slot[121] EnemyType, for example, the concept for treating a

---

[118] D. B. Lenat and E. A. Feigenbaum, OTK, p. 198. For a more detailed discussion of this point, the two AI experts refer to G. Lakoff and M. Johnson, *Metaphors We Live By*, Chicago, IL: University of Chicago Press, 1980.

[119] Cfr. D. B. Lenat and E. A. Feigenbaum, OTK, p. 197.

[120] D. B. Lenat and R. V. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project*, p. 8.

[121] A "Slot" is the technical term for an attribute of an object or event. (Cfr. E. Rich and K. Knight. *Artificial Intelligence*, New York: McGraw-Hill, 1991, pp. 257-275) In this example the events are: treating a bacterial infection and fighting a war.

bacterial infection (Disease) and for fighting a war (MilitaryForce) are analogous. Similarly, in the slot EnemyLocal, the concepts "Bacteria" and "EnemyTroops" are analogous. Not all the concepts in corresponding slots, however, are similar. Lenat and Guha point out that the corresponding concepts that currently fill the slot *Tactics* are not analogous. "Vaccination" and "MilContainment" are not analogous. It is precisely this asymmetry which enables new concepts to be discovered. They affirm: 'Such asymmetry is bound to happen, and this is an opportunity: let the analogy guide the search for new, useful concepts on each side.'[122] Building on the example, they illustrate how this can be brought about as follows:

> 'For instance, maybe we should define a medical analog of *military containment*, and a military analog of *vaccination*. The former might be a *medical containment* - for example, the use of a tourniquet on a venomous snakebite, or the use of quarantine on a virulent plague. The military analog of vaccination might be *fortifying* or *propagandizing*. To take the analogy even more seriously, the military vaccination might entail letting a small group of enemy soldiers overrun some territory before our friendly forces secure it, as a way of driving home to the local populace just how bad the enemy is.'[123]

Analogizing, we can formulate new concepts for treating bacterial infection (medical containment) and new concepts for fighting a war (fortifying and propagandizing). By means of this persuasive example, Lenat and Guha illustrate the use of analogy as a guide for defining new concepts.

Analogy can also help to elaborate the new concepts. The CYC protagonists illustrate how this occurs as follows:

> 'For instance, what precisely is *medical containment* containing? What does it locally contain? How should one do it? To answer those questions, go to the *military containment* concept, look up the answers there, and map them back using the existing analogy:
>
> **MilContainment**
>
> *usefulTacticIn*: Fighting-a-war
>
> *containedType*: MilitaryForce
>
> *containedLocal*: EnemyTroops
>
> *attributedLimited*: Mobility
>
> *howTo*: Bound and Isolate
>
> *counterTactic*: (Threaten containedArea)
>
> *containedArea*: GeographicRegion

---

[122] D. B. Lenat and R. V. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project,* p. 9.

[123] D. B. Lenat and R. V. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project,* p. 9.

That suggests that medical containment, in the case of treating a bacterial infection, is containing a disease, and, more locally, bacteria. It might be done by surrounding and isolating the infected part of the body.'

By looking up more details of what military containment is about, we can flesh out what medical containment might be about. By means of this example, Lenat and Guha illustrate the use of analogy for fleshing out new concepts.

The Knowledge Principle, the Explicit Knowledge Principle and the Breadth Hypothesis constitute the central tenets of CYC theory of intelligence. They express the global vision of intelligence that underlies CYC. Douglas Lenat, in his usual frank and unaffected manner, likens this vision of intelligence to what he calls a "knowledge pump". He affirms:

'This view is best likened to priming a pump. Visualize your brain as a knowledge pump. Knowledge goes in, gets stored, combined, copied, or whatever: from time to time, you say or write or do things that are, in effect, ways for your brain to emit knowledge. On a good day, the knowledge you give out may be as good or better than the knowledge you put in.

No one expects you to be a productive knowledge pump without training and experience-whether we're talking about playing the piano or tennis, writing a check or a novel, or making a U-turn in a car. Your have to invest some learning-and-teaching time and effort before anyone expects you to be competent at a task, let alone to excel at it.'[124]

Though it may appear simplistic, the simile of the knowledge pump illustrates that Lenat is radically convinced that intelligence is about knowledge. In the same article, Lenat goes on to explain at length that it takes common sense for people to understand each other and that it takes common sense to stay focused and learn.[125] While this may be so, can the common sense it takes to do these things be reduced to massive amounts of cross-referenced general knowledge? Granted that cross-referenced general knowledge is important, is it the radical base of intelligence as Lenat advocates, or are there not more fundamental and radical roots of intelligence? A philosophical discussion is necessary in order to appraise the claims of Lenat that intelligence is rooted in massive amounts of knowledge. In chapter four, I discuss this issue.

2.4. The upper CYC ontology

The first challenge in storing common sense is to determine how the huge quantity of knowledge items, their relations and uses etc., are to be organized inside the machine. A large portion of the work on CYC has been spent thrashing out possible solutions to

---

[124] D. B. Lenat, "From *2001* to 2001: Common Sense and the Mind of HAL", p. 195.

[125] Cfr. D. B. Lenat, "From *2001* to 2001: Common Sense and the Mind of HAL", pp. 196-201.

this question and trying to figure out ways to go about realizing such a task in the simplest and most efficient way possible. Lenat and his fellow researchers came up with a scheme which they refer to as the "CYC ontology".

The ontological scheme that was chosen for CYC underwent several phases. It was reworked several times to enhance CYC's ability to resolve common sense problems and to boost its efficiency. In more recent years important changes have been made to cater for contexts. While contexts radically affect the way information is represented in CYC, a level of what may be called context free representation is maintained. The ontological scheme adopted in CYC at this level is referred to as the "upper CYC ontology". A detailed and easy to use guide of the upper CYC ontology can be found at the CYC Web Site.[126] The ontology guide comprises a cover page, a brief introduction, and a table of contents providing links to 43 major groupings of CYC terms. Each time a term appears a link is provided to a brief explanation of that term. Links are also provided to an alphabetical listing of CYC terms when needed.

The public release in Internet of the upper CYC ontology contains approximately 3,000 terms. The CYC team considers these terms to be the core of their ontological scheme. Introducing the upper CYC ontology, Lenat and his colleagues explain:

> 'It is the core of an ontology that satisfies these two important criteria:
>
> It is "**universal**": Every concept one can imagine can be correctly linked into the Upper CYC Ontology in appropriate places, no matter how general or specific, no matter how arcane or prosaic, no matter what the context (nationality, age, native language, epoch, childhood experiences, current goals, etc.).
>
> It is "**articulate**": The distinctions which are made in the ontology are both necessary and sufficient for most purposes: By "necessary" we mean that the distinctions are all worth making. There are both theoretical and pragmatic justifications for every collection, for every predicate and function, for every individual. By "sufficient" we mean that enough distinctions have been made to enable and support knowledge sharing, natural language disambiguation, data base cleaning and integration, and other applications.'[127]

Universality and articulation are strong claims. Lenat and his team explain that it is the result of person-centuries of effort expended over the past dozen years. Their selection of upper level concepts has been extensively tested using 'tens of millions of examples.'[128] They encourage researchers in the various AI fields to employ ontologies that can be easily mapped into the CYC upper ontology. This will facilitate the sharing of knowledge between systems. Once other systems share the same upper ontology with

---

[126] Cfr. "Welcome to the Upper CYC Ontology" at http://www.cyc.com/cyc-2-1/cover.html, Cycorp, Inc., Austin TX, 1997.

[127] "Welcome to the Upper CYC Ontology", p. 1.

[128] "Welcome to the Upper CYC Ontology", p. 1.

CYC, differences at lower levels is not an obstacle to sharing knowledge. At lower levels each system can have its more specialized terms according to its scope and the purposes for which it will be used.

Systems that share the CYC upper ontology will also be able to benefit from the vast quantity of lower level terms, notions and rules of thumb, that have been entered into CYC. The definitions of the lower level terms take into account the many uses that a term can have according to the contexts in which it is used. A further advantage lies in the possibility of using CYC to facilitate the exchange of information among applications that share the upper CYC ontology. Lenat and his crew see CYC as a common ground for bringing together applications in diverse areas such as natural language understanding and generation, database integration, semantic information retrieval, ontology-constrained simulation and so on.

2.4.1. Basic concepts

The ontology of CYC is organized around the concept of collections. Each collection is a concept representing a set or class of things that have some properties in common. Guha explains that they are akin to what Quine termed "natural kinds".[129] Collections are like mathematical sets in so far as they may have elements, subsets, and supersets, and may not have parts or spatial or temporal properties. They differ form mathematical sets, however in several important ways. The CYC team explains:

> 'Sets differ from collections in that a mathematical set may be an arbitrary set of things which have nothing in common. In contrast, the elements of a collection will all have in common some feature(s), some intensional qualities.'[130]

Given that collections are distinguished by their features or their intensional qualities, two collections can be co-extensional (i.e., have all the same elements) without being identical. In contrast, two mathematical sets with the same elements are considered identical.

In CYC the relation of one term being a subset respect to another is expressed using the predicate *genls*. Thus x *genls* y is the same thing as x is a subset of y. *genls* is also used to express the relation that one term is a superset of another. Thus x *genls* y means that y is a superset of x. To express the relation that one term is an element or instance of a collection, CYC uses the predicate *isa*. Thus x *isa* y is the same thing as x is an element of y, or x is an instance of y. The two predicates *genls* and *isa* illustrate the fact that subsets and elements are sharply distinguished in CYC. The upper ontology caters for the fact, for example, that the relation between JohnMcCarthy and Person is

---

[129] R. V. Guha and D. B. Lenat, "CYC: A Midterm Report", *AI Magazine* Fall (1990), p. 42.

[130] "CYC Fundamental Vocabulary" at http://www.cyc.com/cyc-2-1/vocab/fundamental-vocab.html, Cycorp, Inc., Austin TX, 1997, p. 2.

different from the relation between ComputerScientist and Person. JohnMcCarthy is an element of the set Person whereas ComputerScientist is a subset of the collection Person.

The universal set is $Thing.[131] It is the collection of everything. Every CYC constant in the knowledge base is an element of this collection and every collection in the knowledge base is a subset of the collection $Thing. An important partitioning of $Thing is into the two sets $Collection and $Individual. $Collection is the collection of all collections. $Individual is the collection of all things that are not sets or collections. $Individual includes many types of objects. They may be physical objects, temporal objects, numbers, relations, groups and many other kinds of things. Objects that have parts can be considered as a collection of its parts. The CYC team emphasizes, however, that an individual that has parts is distinct from a collection containing those same parts. They write:

> 'Though an element of $Individual may have parts (e.g., $physicalParts or $groupMembers), that individual is NOT the same thing as the collection containing those same parts. For example, your car is an individual, but the collection of all the parts of your car is an instance of $Collection. The latter -- the collection of parts of your car -- is an abstract collection; it does not have a location, it does not have a top speed, etc. -- it is just a collection! -- but it does have subsets, super sets, and members. Similarly: "Bill Clinton's immediate family" is an individual; however, the collection of persons who belong to that family is a collection. One final example: A company belongs to $Individual and is distinct from the collection of its employees (which $isa $Collection).'[132]

Individual objects are always non-sets and this is their defining characteristic. $Individual and $Collection are disjoint collections. No CYC constant can be an instance of both.

Together with $Collection and $Individual, another basic concept of the CYC ontology is the term $Predicate. It is the set of all CYC predicates. The elements of this collection are truth-functional relationships which take some number of arguments. They always return either true or false. Relationships involving opaque contexts such as modal contexts and propositional attitudes are not included in $Predicate. Quantifiers and logical operators are also excluded. The naming strings for elements of $Predicate always begin with a lowercase letter. Two of the most important predicates are $isa and $genls which were described above.

---

[131] To facilitate the discussion in this section, I use the symbol "$" as a prefix to distinguish CYC terms from words that are part of the descriptive text. If the first letter in a CYC term is a capital letter, then the term represents a collection. When the first letter is a common letter, the concept is a predicate. Thus $Collection is a collection and $isa is a predicate.

[132] "CYC Fundamental Vocabulary", p. 1.

2.4.2. CYC fundamental vocabulary

Together with the basic concepts described above, Lenat and his colleagues have established several other fundamental terms for organizing common sense knowledge. A series of collections and predicates, for example, handle different types of disjoint collections and the exceptions that can arise to their normal application.[133] Another group of collections handle relationships that cannot be reduced to true-false functions. Examples of such relations are mathematical and statistical functions, units of measure, recursive relations, modal relations and propositional attitudes.[134] Collections are incorporated in the CYC fundamental vocabulary to describe different types of predicates (unary, binary, ternary, transitive, intransitive etc.) and different kinds of functions (mathematical, unit of measure, collection denoting, non predicate etc.) and to indicate the types and formats of arguments that a predicate can use. The fundamental vocabulary also includes logical connectives such as $implies, $not, $and, $or, $forAll, $thereExists, $different and several others which all serve to connect terms. There are also collections for the different kinds of expressions in CYC. Three types of expressions for example, are linguistic, computational and string expressions.

To maintain consistency in the upper ontology, CYC fundamental vocabulary includes predicates for default reasoning such as $overrides and $exceptfor. $overrides is used to tell CYC which rule to prefer when it encounters two conflicting rules while reasoning with default assertions, and $exceptfor is used to state exceptions to rules. In the fundamental vocabulary, collections are included for the different types of attributes that a thing can have (composite, ordered, unordered, feelings and shape for example) and for the sets of values that an attribute can possess (weather, orientation, physical attribute etc.). Predicates are provided to relate attribute types and values. Finally, the fundamental vocabulary includes a series of specialized collections and predicates to keep track of CYC code and to monitor the development of the database.

2.4.3. CYC top level vocabulary

Certain common sense realities have proven particularly challenging to represent. In this section some of these realities will be discussed, describing the particular solution adopted in CYC.

The first issue concerns the makeup of the universe. It deals with the long-standing philosophical problem of substances and individuals. What are individuals made of? What is everything made of? How do individuals relate to what they are made of? The CYC collection $StuffType captures the notion of substance in CYC. Lenat and his team explain this term as follows:

---

[133] Cfr. “CYC Fundamental Vocabulary”, pp. 3-6.

[134] “CYC Fundamental Vocabulary”, pp. 6-8.

'A collection of collections. Every element of $StuffType is a collection of substances which have the following logical property: such a substance may be subdivided, spatially or temporally, and the resultant portions will also be instances of the $StuffType collection to which the original substance belonged. Elements of $StuffType may be collections of any kind of stuff, tangible or intangible, temporal or atemporal, which has that property (The notion of $StuffType corresponds roughly to that of a mass noun in English.)'[135]

The notion of substance in CYC is classified as a collection. In classical philosophy, a substance is a metaphysical reality which makes something what it is and not another. Substance in CYC is not to be confused with the classical notion. CYC considers substances to be sets of things and not metaphysical entities. Substances must satisfy, however, a determined condition. They can be subdivided into elements and each of the elements is of the same substance from which it came. Clear examples of substances are water, sand, air and butter. Not so obvious examples include time, walking and breathing.

CYC demands that all individuals be made of some substance or the other. Conversely, every piece of any substance is clearly an individual. This leads to the result that substance or stuff is coextensional with the collection of individual objects. While this is true, Guha explains that they still choose to distinguish between stuff and individuals. The two collections may be extensionally equivalent but they have different intensional descriptions. An important difference in the intensional plane concerns the notion of intrinsicness and extrinsicness. Guha writes:

'The notion of intrinsicness is closely related to that of substances: Consider a particular table made entirely of wood. It inherits various default properties from $Wood, which is the kind of substance it is an instance of (density, flashpoint etc.), and it inherits other properties from $Table, which is the kind of individual object that it is an instance of (number of legs, cost, size, etc.). The former properties are intrinsic, the latter are extrinsic.'[136]

As the example illustrates, an object typically inherits its intrinsic properties from the type of substance of which it is an instance, and it inherits extrinsic properties from the type of object that it is an instance of.

An important subset of the collection $StuffType is the subset $TemporalStuffType. This subset expresses the fact that aside from spatial divisions, divisions can be temporal intervals or slices. The class $AnimalWalkingProcess, for example, has the property that when a member of the class is temporally carved into pieces, each one is still an instance of $AnimalWalkingProcess. $AnimalWalkingProcess is thus a type of temporal substance. It is an instance of $TemporalStuffType.[137] Although $Animal-

---

[135] "CYC Top-Level Vocabulary" at http://www.cyc.com/cyc-2-1/vocab/top-vocab.html, p. 2.

[136] R. V. Guha and D. B. Lenat, "CYC: A Midterm Report", p. 44.

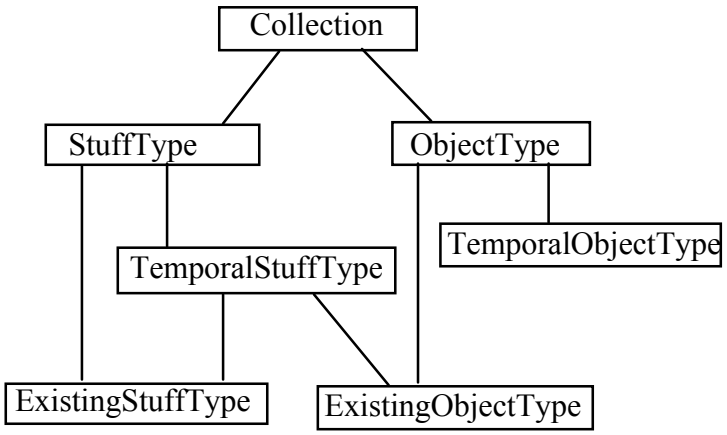
[137] Cfr. "CYC Top-Level Vocabulary", p. 3.

WalkingProcess is a type of temporal substance, $WalkingToTheMailboxAndBack is not. Guha explains:

> 'If you imagine the third minute of the 10-minute $WalkingToTheMailbox-AndBack event, it is still an instance of $Walking, but a stranger watching just this minute would not say that it was an instance of someone walking to a mailbox and back home.'[138]

The collection of all the $WalkingToTheMailboxAndBack events is not an instance of $Walking. It is rather a temporal object, an instance of $TemporalObjectType. The relationship between $Walking and $WalkingToTheMailboxAndBack is indeed the same as the one between $Wood and $Table in the example above where the relationship between a substance and an individual made of that substance was explained. Thus an instance of the event $WalkingToTheMailboxAndBack is an instance of $Walking from which it inherits default values for rate of speed, step size etc. From the event $WalkingToTheMailboxAndBack, the instance inherits default values for destination, duration and so on.

$TemporalStuffType and $TemporalObjectType are important subsets of $StuffType and $ObjectType respectively. They capture the temporal reality of things. Other subsets of $StuffType and $ObjectType capture the spatial reality of things. The separation of temporal and spatial dimensions is an important characteristic of common sense reasoning in CYC. In common sense reasoning, however, the notion of something often includes a reference to the fact that it is something that exists. There are implicit temporal references. To cater for this reality, the CYC team include two important collections in their top level vocabulary: $ExistingObjectType and $ExistingStuffType. Most tangible objects are elements of $ExistingObjectType. Notions such as $Wood, $IceCream and $ApplePie are elements of $ExistingStuffType.[139] The following diagram illustrates the relationship among the terms that have been discussed to this point:



[138] R. V. Guha and D. B. Lenat, "CYC: A Midterm Report", p. 44.

[139] Cfr. "CYC Top-Level Vocabulary", pp. 3-8.

Figure 2.1. $StuffType, $ObjectType and related constants

Related to the notion of substance in CYC is the notion of "granule". A granule is the building block out of which a substance is built. It was explained above that substances must satisfy a determined condition. They can be subdivided into elements and each of the elements is of the same substance from which it came. Division, however, cannot go on indefinitely. Describing the predicate $granuleOfStuff, the ontology release explains:

> 'Eventually, the division of something spatially stuff-like will result in the object-like "granules" out of which the stuff-like thing is composed. For instance, division of sand would eventually result in individual grains of sand, division of water would eventually get down to individual molecules of water, etc. At this level of division or below, the remaining physical portions do NOT count as instances of the stuff-type from which they were divided.'[140]

In Lenat's ontological scheme, an individual molecule of water does not count as water. This is so because it does not have most of the properties that the group made up of many molecules of water has. It is not even stuff-like. In the CYC ontology, individual granules do not count as instances of the collection of which they are granules.

For $TemporalStuffType, the granule predicate $granuleOfTime is used to describe the granule of time associated. The underlying notion is the same as with $granuleOf-Stuff. Divisions cannot go on indefinitely. Describing the term, the CYC team explain:

> 'Eventually, the division of something temporally stuff-like will result in the temporal object-like "granules" out of which the stuff-like thing is composed. For instance, division of a walking process would eventually result in individual steps. At this level of division or below, the remaining temporal slices do NOT count as instances of the temporal stuff-type from which they were divided.'[141]

In CYC's ontological scheme, an individual step does not count as a walking process. This is so because it does not have most of the properties that the group made up of many individual steps has. It is not even temporally stuff-like.

The collection $TemporalThing in CYC, gathers together the collection of all things which have a particular temporal extent. It is an element of $TemporalStuffType and a subset of $Individual and includes elements that are defined solely by their temporal properties such as "the year 1849", "this instant", and "the 21st century". Some things are not instances of $TemporalThing because they are abstract or timeless. These are notions such as mathematical sets, integers, attributes and so on. $SomethingExisting is the subset of $TemporalThing whose elements are more or less static, at least compared to the highly dynamic elements of $Event. According to the upper CYC ontology release, the clearest examples of $SomethingExisting are tangible things such as people,

---

[140] "CYC Top-Level Vocabulary", p. 17.

[141] "CYC Top-Level Vocabulary", p. 16.

lakes, stars, the earth's ionosphere, etc. The document explains that some intangible things such as agreements and obligations exist stably in time over their lifetime rather than merely happening and are thus subsets of $SomethingExisting.[142]

The top-level CYC vocabulary includes several important subclasses of $SomethingExisting. Two of these are $PartiallyTangible and $PartiallyIntangible. The public release describes the concept of partially tangible as follows:

> 'Elements of $PartiallyTangible have some tangible (i.e. material) part and also have a temporal extent (i.e. they exist in time). They may or may not also have an intangible part; e.g., a book is made of matter has a temporal extent, and also has intangible content which is the information content of the text that the author wrote.'[143]

The class $PartiallyTangible is an element of $ObjectType and of $ExistingStuff-Type. The class $PartiallyIntangible is the collection of things having an intangible component but which exist in time. Some intangible things such as the number 2000 have no temporal aspect and thus the class of Intangible things is not a subset of $PartiallyIntangible. Some intangible things, however, such as the laws of a particular country do exist in time and are elements of $PartiallyIntangible.

The classes $PartiallyTangible and $PartiallyIntangible have two important subsets. These are $TangibleThing and $IntangibleExistingThing respectively. $TangibleThing is defined in CYC as the collection of things which are made of some sort of matter and whose nature is primarily material in the sense that they do not have important non-physical properties such as encoded information. $IntangibleExistingThing is the set of things which are intangible yet exist in time. It is the set of things such as codes of conduct, standards for acceptance, bank accounts and so on.

The concepts $PartiallyTangible, $PartiallyIntangible and their subsets are temporal things which have a stable existence over their lifetime. Some temporal things are events or actions which do not display a stable existence but are rather things that happen. They capture changes in the state of the world. The class of such temporal things is the collection $Event. The CYC team explain that the term includes both physical events, such as a rock falling or a bird flying, and mental actions, such as a person thinking or learning. $Event is also a subset of $Intangible since an event consists of the actions per se which is intangible. The actions, in turn, refer to the tangible objects which participate in them. Some temporal things are neither stable existing things nor events that happen. Such is the case with pure disembodied elements such as "the year 2000", "this instant", and "the 21st century". The CYC term $Time-Interval gathers together such concepts. The following diagram illustrates the relations between the important CYC terms that have been discussed above.

---

[142] Cfr. "CYC Top-Level Vocabulary", p. 5.

[143] "CYC Top-Level Vocabulary", p. 5.

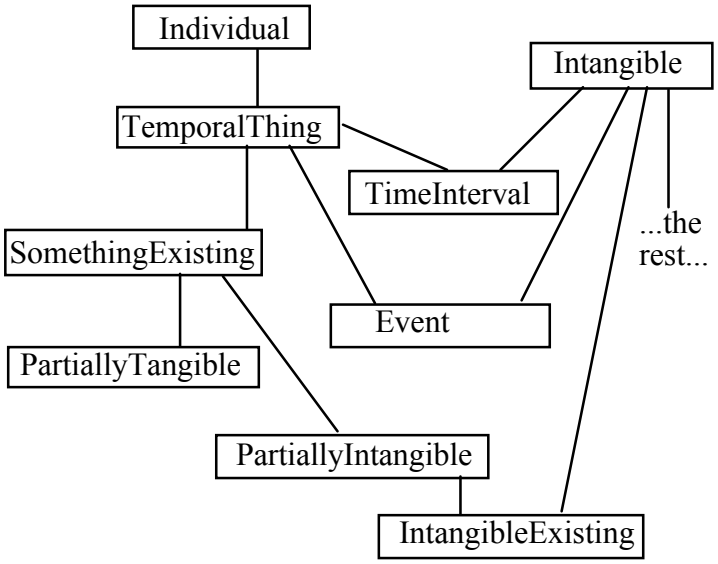


Figure 2.2. $TemporalThing and related constants

A series of predicates capture important aspects of events. The predicate ($subEvents WHOLE PART) expresses the reality that the event PART is a meaningful part of the event WHOLE. It can be used to decompose events into sub-tasks. The predicate $cotemporalSubEvents is used to relate an event to some sub-portion of the event which has the same duration as the whole event but does not include everything that happens. $postEvents relates one event to another that follows it. There ought to be some significant relationship (causal for example) between the two. There are predicates to describe the configuration of things or the situation of things before and after an event occurs. The top-level vocabulary includes several important subcategories of events. The term $PhysicalEvent describes events which involve the interaction of some number of physical objects. $CreationOrDestructionEvent is a physical event in which one or more of the entities involved come into or go out of existence. Two of its subsets are $CreationEvent and $DestructionEvent. Some events are both creative and destructive. The CYC term $TransformationEvent is a collection of events in which at least one thing ceases to exist and at least one thing comes into existence. The ontology guide explains that at least one portion of the thing(s) destroyed becomes incorporated into the thing(s) that are created.

A top-level type of event is the collection $IntrinsicStateChangeEvent. It is the collection of events which are characterized primarily by a change in some intrinsic property of one main entity involved in the event. The CYC team explain:

> 'Such intrinsic changes may include changes of a thing's color, temperature, device state, size, and so on. Events where the main change is extrinsic (such as a change in location or ownership) are not $IntrinsicStateChangeEvents. In events which have more than one actor, the event may be an $IntrinsicState-ChangeEvent for one actor but not for another. For example, in a $Fastening-SafetyBeltEvent, the $SafetySeatBelt (the device used) goes from unconnected to connected (to itself), which is an intrinsic change; however, the agent who

does the fastening (i.e.., who is $HandlingADevice) does not change intrinsically, but only in its configuration to the belt, an external object.'[144]

Some subsets of $IntrinsicStateChangeEvent are: $PreservingFood, $PreparingFood, $ShapingSomething, $BiologicalDevelopment, $GoingToSleep, $SimpleRepairing, and $RecoveringFromAilment.

Another important type of event is captured using the term $GeneralizedTransfer. Each element of this collections is a general kind of transfer event in which something, tangible or intangible, is transferred from one location to another. $GeneralizedTransfer includes changes in physical location, in ownership or possession, transfer of information, and propagation of wave phenomenon through space. $MovementEvent is a subset of $GeneralizedTransfer which is the most general collection of events that are physical movements. The CYC upper ontology explains:

'Each element is an action primarily about some element(s) of $PartiallyTangible rotating or translating, periodically or nonperiodically, with respect to some frame of reference which is not a part of the $objectMoving.'[145]

The guide notes that a person raising her hand and waving, or a tree whose branches are bending in a strong wind do not move in the sense of $MovementEvent. This is so because they remain in the same place. The hand and the branches, however do move. The person and the tree have $subEvents which belong to $MovementEvent.

Among the major types of events is the CYC collection $Action. Each instance of $Action is an event that is carried out by some doer. The doer may be one or more actors who effect some changes in the tangible or intangible state of the world. Normally there is some expenditure of effort or energy. The CYC team explains that it is not required that any tangible object be moved, changed, produced, or destroyed for an action to occur. The effects of actions may be intangible. Depending on the context, actors may be animate, inanimate, conscious or nonconscious. Important types of action are those that are intentional.

The aim of this section was to highlight the major terms that makeup the upper CYC ontology. The discussion centered around the high level basic concepts and the CYC top level vocabulary. The notions, divisions and relations at this level set the framework for describing new terms and for reasoning in a common sense manner. They are the core of the CYC ontology and, as a result, of all the CYC terms, they are the most relevant for the discussion in the following chapters. The description offered is incomplete but illustrates well the style and scope of CYC.

---

[144] "CYC Top-Level Vocabulary", p. 10.

[145] "CYC Top-Level Vocabulary", p. 11.

2.5. How CYC has progressed

Throughout the history of CYC, Lenat and his crew have remained faithful to their three guiding principles: the Knowledge Principle, the Explicit Knowledge Principle and the Breadth Hypothesis. As the CYC knowledge base grew, however, some techniques proved unworkable and were abandoned in favor of new methods. In this section I briefly describe the history of this process.

2.5.1. The three stages of CYC development

In his characteristic unassuming outspoken style, Douglas Lenat summarizes the three stages for developing CYC as follows:

> '1. Prime the pump with the millions of everyday terms, concepts, facts, and rules of thumb that comprise human consensus reality - that is, common sense.
>
> 2. On top of this base, construct the ability to communicate in a natural language, such as English. Let the HAL-to-be use that ability to vastly enlarge its knowledge base.
>
> 3. Eventually, as it reaches the frontier of human knowledge in some area, there will be no one left to talk to about it, so it will need to perform experiments to make further headway in that area.'[146]

With the metaphor "prime the pump" in the first stage, Lenat likens the human brain to a knowledge pump. Knowledge goes in, it is moved around, and there is an output of knowledge from time to time when the subject says, writes or does something. Lenat is convinced that a broad base of knowledge is necessary before a system can rapidly assimilate new knowledge items in a way that is not merely superficial. This primitive knowledge base must be hand crafted. Advanced techniques for acquiring knowledge such as natural language understanding and machine learning are not useful at the outset. This is so because the system lacks the elementary knowledge that is needed in order to assimilate the information that is presented to it through advanced techniques.[147] A broad base of knowledge must be manually constructed in order to support more advanced learning techniques later on. Hand crafting this base of knowledge is what Lenat means by "prime the pump".

The hand built knowledge base links together the millions of everyday terms, concepts, facts and rules of thumb that comprise human consensus reality or common sense. As the knowledge base grows, a point is reached where the system has sufficient knowledge to support natural language understanding. The system can assimilate the information that is presented to it by means of natural language in a way that is not

---

[146] D. B. Lenat, "From *2001* to 2001: Common Sense and the Mind of HAL", pp. 201-202.

[147] Cfr. D. B. Lenat and R. V. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project,* pp. 24-27.

merely trivial but rather at a high level of competence, interpreting the information in function of what it already knows. At this point stage two can be fully enabled. The system has sufficient common sense to learn through natural language. In the second stage the rate of learning greatly increases. As more is known, the rate of learning continues to grow.[148]

During the second stage, the dominant method for acquiring knowledge is through natural language understanding. As the system acquires more knowledge, it becomes increasingly profitable to favor techniques whereby the machine learns on its own. Lenat and Feigenbaum refer to this as 'learning by discovery'. They explain:

> 'The more one knows, the faster one can discover still more. Once you speak fluently, learning by talking with other people (Language) is more efficient than rediscovery, until you cross the frontier of what humanity already knows, at which point there is no one to tell you the next piece of knowledge.'[149]

The authors suppose that there is a limit to what can be learnt by absorbing all that can be communicated in a natural language. Furthermore, as the systems knowledge becomes more encyclopedic, it is ever more difficult to tell the system something that it does not already know. The third stage is inaugurated at this point. Lenat and Feigenbaum explain what learning by discovery might involve as follows:

> 'Learning by discovery is meant to include not only scientific research (e.g., cancer research), but also the many smaller-scale event in which someone formulates a hypothesis, gathers data to test it, and uses the results to adjust their "theory". That small-scale case can occur in a (good) classroom; or just by driving the same route to work over various different times of the day (and hypothesizing on rush hour patterns). It involves defining new concepts (at least in principle), formulating new heuristics, and even adjusting or changing one's representation of knowledge.'[150]

Having assimilated the greater part of all that can be assimilated through natural language, a knowledge system can acquire new concepts only by discovering them on its own. It does this by formulating and testing hypotheses. The process involves defining new concepts, finding new ways of solving problems, and adjusting its own reasoning and conceptual schemes based on the results of its investigation. It can be supposed that analogical reasoning will play a principle role in defining new concepts and elaborating their content.[151] Knowing all that has ever been known, computers become protagonists of bold new discoveries.

---

[148] Cfr. D. B. Lenat and E. A. Feigenbaum, OTK, p. 209.

[149] D. B. Lenat and E. A. Feigenbaum, OTK, p. 208.

[150] D. B. Lenat and E. A. Feigenbaum, OTK, p. 209.

[151] Cfr. Section 2.1.3.

2.5.2. The early years of CYC

The three-stage research program described above was launched in 1984. When Lenat moved over to the newly formed MCC (Microelectronics and Computer Consortium) in the fall of that year, the idea was that he would assemble a team of dozens of individuals to tackle the first stage over the subsequent decade. Lenat remembers: 'We would "prime the knowledge pump" by hand crafting and spooning CYC with a couple of million important facts and rules of thumb. The goal was to give CYC enough knowledge by the late 1990's to enable it to learn more by means of natural language conversations and reading (stage 2).'[152]

The first problem that Lenat and his team faced was what knowledge to represent. At first they expected encyclopedias to play an important role. Within a few months, however, they realized that the common sense knowledge that they were looking for was not the knowledge that was gathered together in encyclopedias, but rather the knowledge that a reader is assumed to possess in order to understand an encyclopedia article. How could this underlying knowledge be captured? Lenat explains:

> 'If we could not use encyclopedias for their content directly, we could still use their information indirectly. If we take any sentence from an encyclopedia article and think about what the writer assumes the reader already knows about the world, we will have something worth telling CYC. Alternatively, we can take a paragraph and look at the "leaps" from one sentence to the next and think about what the writer assumes the reader will infer "between" the sentences.'[153]

This manner of proceeding, parting from examples and thinking about what the author assumes the reader already knows, characterized, above all, the early years of activity in the CYC project. Lenat and his crew drew on a wide range of examples, from encyclopedias, newspapers, novels, advertisements, and so on.[154] Due to the immensity of the task, it was important that as many persons as possible could work at the same time making updates to the knowledge base. It was quickly discovered that, though unified in their vision for CYC, the CYC crew thought quite differently about many things. Their assessment of the common sense knowledge behind ordinary phrases often differed, giving rise to inconsistencies in the knowledge base. The CYC team resolved

---

[152] D. B. Lenat, "From *2001* to 2001: Common Sense and the Mind of HAL", p. 203.

[153] D. B. Lenat, "From *2001* to 2001: Common Sense and the Mind of HAL", p. 203. Lenat goes on to give an example to illustrate this technique. He writes: 'For instance, back in 1984 our first example read, "Napoleon died on St. Helena. Wellington was greatly saddened." The author expects the reader to infer that Wellington heard about Napoleon's death, that Wellington outlived Napoleon, and so on.'

[154] Cfr. D. B. Lenat, "From *2001* to 2001: Common Sense and the Mind of HAL", p. 204.

this difficulty by putting into place a series of procedures for detecting inconsistencies before updating the KB. They also established methods by which team members can come to an agreement about the common sense knowledge that ought to be entered when conflicts arise.[155]

Around 1990, as the KB grew, the CYC team opted to engineer the KB working from broader concepts to more specific knowledge items. Lenat and his team invested a great deal of effort in designing this scheme. They refer to it as the Upper or Global Ontology.[156] Reporting on his conversation with Lenat, Lev Grossman summarizes the evolution of the ontological scheme as follows:

> 'Among the categories that emerged was things, followed soon after by subordinate concepts such as tangible objects and intangible objects. The group went on to sketch out events, processes and animals (which are both tangible and intangible at the same time). Soon they came up with stuff, a sub category covering things like peanut butter, which are continuos and homogeneous, as opposed to things like marbles, which are discreet objects.'[157]

Choosing to proceed in this way, the CYC team oriented their efforts towards working in a top-down fashion, treating one topic at a time and in moderate detail. By 1996, they had told CYC about hundreds of topics.[158]

2.5.3. Reworking the representation language

Along with the effort to choose the concepts to represent, the CYC team developed a representation language for manipulating the knowledge elements in the CYC system. They called it the CYC language or CycL.[159] CycL has two parts which are characteristic of representational languages in general: on one hand it defines a method for defining the elements of the KB and how they are related, and on the other hand, it incorporates a method or language for deriving conclusions from the KB. Procedures for deriving conclusions from the KB are termed *inference* procedures.

---

[155] Cfr. D. B. Lenat and R. V. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project,* pp. 33-35.

[156] Cfr. section 2.4.

[157] L. Grossman, "Get Smart: How Intelligent Can Artificial Intelligence Be? Lev Grossman Talks To Doug Lenat To Find Out", *Time-Digital Magazine* (October 1998).

[158] Cfr. D. B. Lenat, "From *2001* to 2001: Common Sense and the Mind of HAL", p. 204.

[159] Cfr. D. B. Lenat and R. V. Guha, "Enabling Agents to Work Together", *Communications of the* ACM 37 (1994),p. 128.

Initially, a simple frame-and-slot language was used to store information in CYC.[160] Lenat and Guha recall:

> 'Until 1987, CycL was essentially a set of procedures for performing inheritance, automatic classification, etc., that operated on a set of frame-like data structures. Uncertainty was expressed by associating certainty factors (numbers between 0.000 and 1.000) with every statement in the KB.'[161]

The frame-based representation, however, caused several problems. It was not flexible enough to capture qualifiers such as not, or, every, and some, and to express opinions, expectations, counterfactual conditionals, and other similar realities in an efficient manner.[162] Much of the knowledge that we ordinarily use proved too complex to squeeze them efficiently into the frame-and-slot straightjacket. Lenat and his crew also came to realize that having an implementation-independent semantics for the KB was important. The way knowledge items are defined ought to be independent of the way that CYC goes about deriving conclusions from the KB. In frame-based systems a change in the inference mechanism often demands that the entire KB be redone to support the new way of deriving conclusions. These factors slowly moved the CYC representation language away from frames and towards a predicate logic orientation. This gave it more expressiveness and also provided an implementation independent semantics for the KB.[163]

2.5.4. Three important lessons

Alongside the move from frame representation to predicate calculus logic, Douglas Lenat highlights three principle lessons that have been learned along the way. As mentioned above, uncertainty was originally expressed by associating certainty factors (numbers between 0.000 and 1.000) with every statement in the KB. Uncertainty factors provided a means by which statements could be compared in order to determine which is preferable. This proved too simplistic a method. Lenat writes:

---

[160] A frame is a collection of attributes (usually called slots) and associated values (together with constraints on those values if they apply) that describes an entity in some absolute sense (Cfr. E. Rich and K. Knight, *Artificial Intelligence*, pp. 257-275). A single frame taken alone is rarely useful. Frames are connected to each other by virtue of the fact that the value of an attribute of one frame may be another frame. Connected together, frames structures can support powerful reasoning procedures.

[161] D. B. Lenat and R. V. Guha, "Enabling Agents to Work Together", p. 128.

[162] Cfr. D. B. Lenat, "From *2001* to 2001: Common Sense and the Mind of HAL", p. 204.

[163] Cfr. D. B. Lenat and R. V. Guha, "Enabling Agents to Work Together", p. 128.

> 'Including certainty factors had several bad consequences, however, and we eventually changed to a scheme in which all inputs are true by default. To decide whether to believe something, CYC gathers up all the pro and con arguments it can think of, examines them, and then reaches a conclusion.'[164]

This procedure in CYC is called *argumentation* and it is one of the key inference methods employed in CYC.[165]

The second lesson concerns the tradeoff in every representational language between expressiveness (how easily you can say complicated things) and efficiency (how easily the machine can reason with what it has been told). Natural languages such as English and Spanish are very expressive but not very efficient. Most computer languages, on the other hand (Basic, C, FORTRAN, etc.) are efficient but not very expressive. The challenge is to find the balance that best meets the demands of the system in question. In the case of CYC, expressiveness is needed on the user end to facilitate the interaction between the users of CYC and the system. For its internal inference procedures, an expressive language is important to the extent that the programming language used ought to be minimally intelligible. Otherwise it would be impossible to program CYC. For internal processes expressiveness can be sacrificed to gain efficiency. In order to obtain the benefits of expressiveness and efficiency at the same time, Lenat and his colleagues developed two separate languages for CYC. Lenat explains:

> 'To get both qualities, we separated the epistemological problem (what the system should know?) from the heuristic problem (how can it effectively reason with what it knows?) and developed two separate languages, respectively EL and HL. Our knowledge enterers talk to CYC in the clean, expressive language (EL). Their input is then converted into the heuristic language (HL), which is efficient for dealing with many sorts of frequently recurring inference problems, such as reasoning about time, causality, containment, and so forth.'[166]

Separating the epistemological problem (what the system should know?) from the heuristic problem (how can it effectively reason with what it knows?) Lenat and his team were able to boost efficiency while maintaining expressiveness where it is most needed.

The third lesson concerns an issue which has been of central concern over the last four years - incorporating contexts. Lenat and his team originally envisioned a single large knowledge base of consensus reality that would depend little on contexts. As the

---

[164] D. B. Lenat, "From *2001* to 2001: Common Sense and the Mind of HAL", p. 205.

[165] The details of the argumentation procedure is described by Guha in: R. V. Guha, "The Representation of Defaults in CYC", *MCC Technical Report* CYC-083-90, Austin, TX: Microelectronics and Computer Consortium, 1990.

[166] D. B. Lenat, "From *2001* to 2001: Common Sense and the Mind of HAL", p. 205.

KB grew however, it became increasingly difficult to maintain consistency. Looking ahead around 1990, the group realized that heroic efforts would be required to explicitly define the valid and invalid contexts of use, each time CYC is updated with a new expression. They also realized that the assumptions that are latent in a given context can reduce the vocabulary that is needed to convey meaning within that context. A context independent KB will not be able to take advantage of this benefit. Each expression must be entered by means of an expressive cumbersome vocabulary. Faced with the heroic efforts that avoiding contexts entailed, Lenat and Guha concluded:

> 'Even after these heroic efforts, the knowledge in the KBS will generally only be partially decontextualized, especially if it is about people doing things in the day-to-day common-sense world. Attempts at producing completely decontextualized representations, in such domains, are destined to be futile.'[167]

Incorporating contexts has perhaps been the most significant change in CYC since its conception in 1984. Lenat writes: 'Perhaps the most important lesson we learned along the way was that it was foolhardy to try to maintain consistency in one huge flat CYC knowledge base.'[168] Incorporating contexts meant that the CYC KB had to be reconstructed altogether. Describing this change, Lenat explains:

> 'We eventually carved it (the KB) up into hundreds of contexts or microtheories. Each one of those is consistent with itself, but there can be contradictions among them... In the fictional context of Bram Stoker's *Dracula*, vampires exist; in the standard rational world view context they don't. Other contexts carve out similar distinguishable eras in time, political or religious points of view, and so forth.'[169]

The CYC representational language has been modified so that assertions are entered together with their context. Assertions in CYC are not universally true unless they pertain to the universal context. They are true or false depending on the context in which they are affirmed. How contexts are represented in CYC is discussed in the following chapter.

2.5.5. Applications of CYC

One of the spectacular early uses of CYC has been for information retrieval. In a current prototype for image retrieval, CYC is given an on-line database of digitized still images and videos, each of which has a caption. The captions are entered in English and a CYC operator translates them into CycL axioms. Similarly, querying is done by

---

[167] D. B. Lenat and R. V. Guha, "Enabling Agents to Work Together", p. 134.

[168] D. B. Lenat, "From *2001* to 2001: Common Sense and the Mind of HAL", p. 205.

[169] D. B. Lenat, "From *2001* to 2001: Common Sense and the Mind of HAL", p. 206.

issuing CycL queries and CYC performs matching between the captions and the images, carrying out inference when necessary.[170] CYC matched the query "A happy person" against "A man watching his daughter take her first step". Similarly, CYC matched the query "a strong and adventurous person" to a caption of "a man climbing a rock face." Lenat explains that CYC was able to do the latter using rules like "If people do something for recreation that puts them at risk of bodily harm, then they are adventurous."[171]

In another test, the CYC system is queried for images evoking cuteness. It first asks whether one means cute as in attractive or cute as in heart-warming. After the latter is chosen, it finds many such images. A typical one is captioned "A dog being carried in a backpack." Lenat and Guha write: 'Note that no element of the picture - dog, carrying, backpack - is inherently cute, but CYC knows that animals in unusual places are cute.'[172] Weighing in the evidence from their image retrieval prototype, Lenat and Guha see a confirmation of their knowledge based approach to AI. They conclude:

> 'This semantic image retrieval prototype illustrates how relatively shallow knowledge about a very large spectrum of topics can change the way information retrieval works.'[173]

Another prototype CYC application is for understanding structured information sources such as spreadsheets and databases, and then using that understanding to detect common sense errors and inconsistencies in the data. Lenat claims that a fair degree of success has already been obtained in this area.[174] Applications of CYC in information analysis are already on the market. Cycorp's clients include Glaxo, Wellcome and Kanisa. These companies use CYC to perform tasks that are repetitive and tedious but require a certain degree of common sense knowledge. One product in development would employ CYC to seek out common sense errors in databases. If twins are listed with different ages, for example, CYC detects that something is wrong.[175]

---

[170] Cfr. D. B. Lenat and R. V. Guha, "Enabling Agents to Work Together", p. 141.

[171] Cfr. D. B. Lenat, "From *2001* to 2001: Common Sense and the Mind of HAL", p. 207.

[172] D. B. Lenat and R. V. Guha, "Enabling Agents to Work Together", p. 141.

[173] Cfr. D. B. Lenat and R. V. Guha, "Enabling Agents to Work Together", p. 141.

[174] Cfr. D. B. Lenat, "From *2001* to 2001: Common Sense and the Mind of HAL", p. 206.

[175] Cfr. L. Grossman. "Get Smart: How Intelligent Can Artificial Intelligence Be? Lev Grossman Talks To Doug Lenat To Find Out".

Despite the progress that has been made, there are important obstacles that have to be overcome. The strategy for representing contexts using microtheories has proven insufficient. In order to resolve the shortcomings that have been encountered, Lenat has proposes a new strategy to represent contexts in CYC. His proposal is described in detail in the third chapter and discussed in chapter four. An additional obstacle concerns the conceptual and technical hurdles that the implementation of the second stage of the project involves. The second stage of the project is aimed at building in the ability to understand a natural language such as English or Spanish. Without such an ability, CYC's knowledge will always be rudimentary. Lenat and Guha explain:

> 'Being able to understand texts is an important part of the CYC long range plan. In the long run the only way that CYC is going to be able to keep abreast of any reasonable fraction of the happenings in the world is by reading texts.'[176]

Work on natural language understanding (NLU) has been underway for over four years but CYC is still unable to understand everyday English. Lenat sees the goal of NLU as still far off. In the short run Lenat and his colleagues have sought to build in semiautomatic tools to accelerate knowledge acquisition. Lenat affirms 'Long before it can read all on its own, CYC will carry on semiautomated knowledge acquisition from texts, a sort of tutoring program in which it asks clarifying questions when it comes across something it is not sure about.'[177] In its current state, CYC can assimilate texts such as almanacs, that are mostly hard data. In a recent test conducted by the Defense Advanced Research Projects Agency, CYC read a set of foreign policy documents, including the CIA *World Fact Book*.[178]

2.5.6. CYC today

Those who are interested in CYC can obtain abundant information about the project by visiting the Cycorp site on the World Wide Web.[179] Cycorp, Inc., as indicated above, was founded by Douglas Lenat and his colleagues in 1995, to develop and market commercial versions of CYC when funding for CYC at MCC ran out. Cycorp, Inc., based in Austin, Texas, promotes itself as the leading supplier of formalized common sense. Lenat is president and CEO of the company. Cycorp offers a family of products to potential customers. The product family comprises an immense multi-contextual knowledge base, an efficient inference engine for reasoning, a set of interface tools, and

---

[176] D. B. Lenat and R. V. Guha, "Enabling Agents to Work Together", p. 137.

[177] D. B. Lenat, "From *2001* to 2001: Common Sense and the Mind of HAL", p. 206.

[178] Cfr. L. Grossman. "Get Smart: How Intelligent Can Artificial Intelligence Be? Lev Grossman Talks To Doug Lenat To Find Out".

[179] Cfr. "Cycorp. Creators of the CYC Knowledge Base" at http://www.cyc.com, Cycorp, Inc., Austin, TX, 1999.

a number of special-purpose application modules. The CYC software runs on Unix and Windows NT which means that it can be installed and run on a desktop PC. The knowledge base is built on a core of over a million hand-entered assertions that are designed to capture a large portion of what is normally considered to be consensus knowledge about the world. It is the foundation of the CYC software and occupies little more that 100MB.[180]

The staff at Cycorp currently consists of approximately 45 members. They constitute a talented group with a broad variety of educational backgrounds. They hope to bring on new members in the coming months as the demand for the product continues to grow. Interested persons can explore employment opportunities at Cycorp by consulting their extensive web site.

To improve and promote its products, Cycorp is currently seeking to form relationships with software publishers, solution providers and system integrators to develop and market the best knowledge-driven solutions in the industry. They have established a partners program to facilitate this initiative. Extensive information about the program is available on their official Web site. In addition to the partners program, Cycorp offers a two pronged support program to provide assistance to their licensed developers and other clients. On one hand clients can call in for rapid assistance during office hours. In addition they can consult an online library for documentation, technical notes and bug reports at any time. Cycorp offers a consultation service to help clients identify and exploit the best ways to integrate Cycorp products into their information technology infrastructure. Furthermore, the company periodically offers training courses throughout the year at the Cycorp headquarters in Austin, Texas. Course outlines can be obtained at the company's web site.

The major challenge facing Lenat and his colleagues at present concerns context representation. The models and techniques that have been developed and tested to date have proven insufficient. Lenat has recently proposed a new strategy to incorporate contexts which pretends to overcome the shortcomings of previous approaches. The evolution of context representation in CYC responds to deeper underlying principles and illustrates how these principles are articulated in the theories and models that are proposed. This process is described in the following chapter.

---

[180] Cfr. L. Grossman. "Get Smart: How Intelligent Can Artificial Intelligence Be? Lev Grossman Talks To Doug Lenat To Find Out".

Chapter III

Context in CYC

In the previous chapter it was mentioned that the most important lesson learnt over the years in building CYC is the need to incorporate contexts. Many of the facts and rules of thumb that make up common sense knowledge have diverse meanings according to the context in which they are used. Lenat and his team at CYC have dedicated a great deal of resources and effort to figure out the best way to cater for contexts in the CYC knowledge base. They do not subscribe to any particular school of thought and the solutions they adopt are ground breaking and original in many ways. Their work sheds light on the scope and possibilities of AI and serves to uncover some of the underlying attitudes and motivations at the roots of AI.

In the first section of this chapter the problem of context in storing common sense is discussed. The original strategy to incorporate contexts in CYC is explained and the shortcomings encountered are discussed. In response to the shortcomings of the original strategy, the conceptual framework for incorporating contexts in CYC has been reworked over the last four years. In the second section I discuss the new conceptual framework that has been adopted and the advances that it promises. In the third section, a detailed discussion of Lenat's new proposal is presented.

3.1. The problem of context in storing common sense

Lenat and his team initially envisioned a flat context-free knowledge base of common sense knowledge. They quickly realized, however, that knowledge items have different meaning depending on the context in which they are used. When people communicate with each other, the phrases and gestures used to transmit information often depend, for example, on what was previously said, on previous conversations and on shared experiences. The person or persons to whom the phrases or gestures are directed interpret what they receive according to the context that they share with the source of information. The Douglas Lenat observes:

> 'If you pluck an isolated line of code from a computer program, or an isolated sentence from a book, or an isolated cell from a spreadsheet, it will likely lose some or all of its meaning. If you show it - "out of context" - to someone else,

they will likely miss some or all of its intended significance.'[181]

Much of the meaning that an object is intended to convey is configured by the context in which the information is interchanged. The part played by the context in communication, however, is a complex reality. Two persons who share a common rich context make references to the context they share in an almost transparent fashion. References are often implicit and tacit and the conversation itself modifies the shared context as it unfolds. While this is so, humans are nevertheless very skilled at using contexts, exercising dominion and control over them, often in an effortless manner. How can contexts be described and modeled so that we can better understand how we use them? How can contexts be stored, updated and used by computers in a manner that is not trivial? These are important interrogatives which Lenat must resolve in order to advance in his project.

3.1.1. Incorporating context in CYC

In 1989, when the CYC KB exceeded 100,000 rules, Lenat and his team found it necessary to introduce an explicit context mechanism. R. V. Guha dedicated his doctoral thesis to study the problem and propose solutions. He led the implementation of a scheme to handle contexts which meant an important breakthrough for the project. The KB was divided into a lattice of hundreds of contexts, placing each assertion in the context to which it belonged.

The first implementation strategy was based on the scheme that for a given assertion, there are certain circumstances in which it applies and others in which it does not apply. The circumstances in which the assertion applies can be represented as a set of assumptions or conditions which serve as tests for applicability. The assertion can be validly applied when the conditions or tests are satisfied. Lenat and his team distinguished between tests that are explicitly determined by the assertion itself and others that are implicit. Tests are explicit when the assertion itself contains a condition. Lenat illustrates this point using as an example, the following assertion: "If it is raining outside, carry an umbrella." A condition or test for applicability is explicitly contained in the assertion. The assertion can be applied subject to the condition that it is raining outside.

Other tests are implicit. They are underlying conditions that cannot be dispensed with but which are normally taken for granted. Lenat suggests that these test can be uncovered by introspection, by having the system get the wrong parses to sentences, or by getting the wrong answer to questions posed to it. Some implicit conditions for the rule: "If it is raining outside, carry an umbrella", would be, for example:

the performer is a human being

---

[181] D. B. Lenat, "The Dimensions of Context Space", Austin, TX: Cycorp, 1998, p. 6. This document will be reffered to hereafter using the abbreviation DCS.

the performer is sane

the performer can carry an umbrella

the performer is not a baby

the performer is not quadriplegic

the performer is going to go outdoors now/soon

the time period is after the invention of the umbrella

etc.[182]

The example above illustrates that each assertion is associated with numerous assumptions. Many of the these are not likely to come to mind right away. The process of introspection by which they are formulated requires effort and time. As a result of this, Lenat and his team understood that it would be too costly to specify the implicit assumptions each time a new assertion was added to CYC. Furthermore, many assertions share the same underlying assumptions and these are not worth stating over and over again. Lenat and his colleagues decided to adopt a more global way to write, store and add to a set of assumptions that are common to a set of assertions. Groups of assertions that share many assumptions were gathered together into CYC objects termed "microtheories". A microtheory together with the assumptions associated with it form a context.

The ability to construct microtheories has been the heart of context representation in CYC. Guha illustrates how they are used as follows:

> 'The CYC language offered the ability to say that a set of sentences constitutes a theory. For example, one set of sentences represents a naive theory of physics (NTP). Here is a typical NTP sentence, which says that if something is not supported, then it falls. (Of course, this heuristic is not always true; it fails for balloons, objects in outer space, and so on. This theory should not be used when dealing with such objects.)
>
> $ist((\forall x \neg supported(x) \subset falls(x)), NTP)$.
>
> These theories, such as NTP, are referred to as microtheories (Mts.). They are first-class objects in the CYC language and are usually given descriptions related to their scope, when they should be used, and so on.'[183]

NTP sentences specify diverse physical regularities in nature that are patent to everyone. The NTP sentences together with their underlying assumptions serve to establish a NTP context. An informal comment is provided to explain what the context is meant to express, in this case, a naive theory of physics.

---

[182] Cfr. D. B. Lenat, DCS, p. 13.

[183] R. V. Guha and D. B. Lenat, "CYC: A Midterm Report", *AI Magazine* Fall (1990), p. 38.

Using the strategy of gathering assertions into contexts, hundreds of contexts have been defined over the past nine years. The use of microtheories has simplified the implementation of contexts in that new assertions can be entered into already existing contexts. When none of the existing contexts capture well the applicability circumstances of a new assertion, relevant assumptions can be copied and edited from existing contexts to form a new context. The technique of importing assertions from one context to another is termed "lifting" in CYC. Using copy and edit functions, new contexts are easier to create and a great deal of introspection is saved.

The strategy of using microtheories also brought benefits for solving problems. Finding solutions to questions are easier when the context is known. To encounter an adequate solution, CYC does not need to search the entire database. The search can be limited to the assertions within the relevant context. Using microtheories thus catalyzes efficient problem solving. Performing inference in this way, however, has the difficulty that the context must be discerned by CYC or indicated by the user. The latter is less desirable as it burdens the user to choose from hundreds of possibilities. Furthermore, using a little common sense it is often clear what the context might be. Lenat explains that heuristic level (HL) modules were employed to recognize certain kinds of questions or statements and handle them using customized, optimized data structures and algorithms that take advantage of the microtheory strategy.[184]

3.1.2. Shortcomings in the original strategy

While microtheories brought important gains, these gains have proven insufficient. The strategy of "lifting" promised to simplify greatly the process of placing assertions into appropriate contexts and to make it easier to create new contexts using copy and edit functions. The experience of the past nine years, however, has shown that the cost of using this scheme is prohibitively high. Lenat explains:

> 'The gains have been undermined and offset by this cost: It is distracting and time-consuming for the knowledge enterer to select (or, even worse, create) precisely the right context each time he/she enters a new piece of domain knowledge. And it can't be skipped: doing a slipshod job of placing each new assertion just leads to bugs (wrong answers, failure to apply relevant knowledge, etc.) and thence a large amount of debugging that must go on - i.e., where a "wrong answer" turns out to be due to someone simply having placed an assertion into the wrong context. Or, just as bad, the knowledge enterer ignores the context mechanism and the result is a small number of enormous contexts.'[185]

Using microtheories has turned out to be more difficult than it was originally hoped. In the first place, the process of lifting proved to be much more complicated than it was at first envisioned. When an assertion is lifted from one context C1 to another C2 to be

---

[184] D. B. Lenat, DCS, p. 15.

[185] D. B. Lenat, DCS, p. 13.

used there in some inference procedure, it ought to be ensured that no logical inconsistency will arise between the assumptions from C1 that are lifted over and the assumptions that already exist in C2. This is important for the viability of CYC. If there are logical inconsistencies among the assumptions associated with a context then the inference mechanism will not work well. The inference mechanism expects logical consistency and relies on this consistency to correctly manipulate information. If logical consistency is not carefully ensured the inference mechanism will produce strange and inconsistent results and thereby lose trustworthiness. Lifting turned out to be a source of logical inconsistencies because when knowledge enterers were not diligent enough, irrelevant assumptions were often copied and assertions were placed in inappropriate contexts.

Lenat explains that in order to maintain logical consistency, the assertions in a context should actually rely on most if not all of the domain assumptions of that context. This provides relief for the knowledge enterers because it reduces the possibility that logical inconsistencies will arise due to irrelevant assumptions being copied during lifting. As a means to ensure logical consistency and to facilitate data entry, Lenat and his colleagues have given a great deal of importance to this point. Procedures were established to correct situations where some of the assertions included in a context do not rely on most or all of the underlying assumptions attached to that context. If an assumption X, for example, was found to pertain to only a small fraction of the assertions in a context then those assertions ought to be split off from the context to form a new context. The assumption X and other relevant assumptions are copied over. The assumption X is then eliminated from the original context. Another way of dealing with this problem, explains Lenat, is to modify the assertions, adding a conditional clause that caters for the assumption X. The assumption X can then be removed. Included as a conditional clause in the assertions where it applies, it has been made explicit and need not be included in the list of implicit assumption. Lenat describes this procedure saying that the assumption X is *conjoined* to the antecedent of each of the problematic assertions.[186]

Lifting, in principle, facilitates data capture and is one of the chief benefits that using microtheories brings. It was particularly useful in the special case where an assertion is imported from a more general context to a more specific context. Such is the case, for example, of importing from a sporting event to a baseball game or other type of sporting activity. In these cases, all the domain assumptions are by default still satisfied in the more specialized context so any assertion can just be imported unchanged from the general context to the more specialized one. Lenat cautions, however, that there are exceptions to the default case and there could be an assertion in the specialized context that contradicts and overrides it. Another special case is that of importing from the most general context in which the assertions are universally

[186] Cfr. D. B. Lenat, DCS, p. 16.

applicable. The assertions in this context are freely importable to every other context.[187]

While lifting was especially useful in certain circumstances, its full fledged use in CYC, however, turned out to be a nagging cause of logical inconsistencies. This stalled data capture because a great deal of time consuming debugging had to be carried out to correct logical inconsistencies. To remedy this difficulty, conjoining was again a useful resort. It provided a means to avoid copying assumptions that would give rise to inconsistencies during lifting. Lenat explains how conjoining enabled this to be achieved with an illustrative example. There are two contexts C1 and C2 as follows: C1 = the context of the Heaven's Gate cult beliefs and C2 = the editorial position of The New York Times in the 1990's. A core assumption A35, for example, of C1 would be: "The Heaven's Gate cult beliefs are right".

Lenat considers the case where an assertion from C1 is lifted into C2. The assertion concerns a claim that the cult makes which The New York Times commented in its editorial during the 1990's. It is the assertion P109, for example, and reads: "The comet Hale-Bopp has a UFO hiding behind it." Assertions such as these, explains Lenat, ought not to be lifted without changes because the underlying assumption A35 that will be brought over ("The Heaven's Gate cult beliefs are right") is surely inconsistent with many of the assumptions that underlie the context of the editorial position of The New York Times in the 1990's. Lenat affirms that asserting P109 in C2 as though it were factual is a mistake that knowledge enterers ought to capture. The assertion that is added to C2 should have a condition conjoined to it to cater for the inconsistency. It may read, for example: "If the Heaven's Gate cult beliefs are right, then the comet Hale-Bopp has a UFO hiding behind it." To summarize, Lenat writes:

> 'In other words, the lifted form of the assertion P109 is not identical to assertion P109 but rather: "If <A35> then <P109>." Assumption A35 was tacked onto the antecedent of P109, because A35 is not one of the assumptions of C2.'[188]

In addition to the problem of irrelevant assumptions, another complicated case in lifting arises when a domain assumption has the form of an implication. A difficulty arises because the lifted form ought to take into account the implications. The information in the implications ought to be captured in the conjuncts of the lifted form. Lenat explains this using an illustrative example in which an assertion is lifted from the context of a military infantry combat. One of the assumptions is that people mentioned anywhere in that context are by default male. The assumptions contain an implication as follows:

```
(domainAssumptions MilInfantryCombatMt
 (implies
```

---

[187] Cfr. D. B. Lenat, DCS, p. 17.

[188] D. B. Lenat, DCS, p. 16.

(isa ?x Person)

(isa ?x MalePerson)))

In the context of military infantry combat, if someone is a person then that person is a male person. Lenat considers the case of an assertion in this context which has a conjunct (isa ?x Person) in its antecedent or its consequent. The lifted form of this assertion into a context which does not assume that people are by default male, affirms Lenat, ought to have that conjunct replaced by (isa ?x MalePerson). To illustrate this point in less technical terms, Lenat writes:

> 'Consider a rule in MilInfantryCombatMt like "almost everyone over 30 has a wife."... Lifted into a less gender-biased military context, it would become "almost every man over 30 has a wife." Lifted into a gender-biased but non-military context, it would become "almost every infantry soldier over 30 has a wife." And if the rule were lifted into a context that was neither gender-biased nor military, then it would become "almost every male infantry soldier over 30 has a wife."'[189]

A related case is where an assertion is accompanied by an implicit assumption that ought to be made explicit when the assertion is lifted to another context. In defining the assumptions for MilInfantryCombatMt, for example, it may have been left implicit the fact that soldiers are persons. Lifted into a context in which some of the performers are not human, a new clause would need to be added stating that the subject is a person. Catering for implicit assumptions in this way is a difficult case because, being implicit, they are not easy to detect. Furthermore, all the assumptions that underlie the knowledge that is entered can never be fully explicated. This, in fact, is an important conclusion of the years of work on context in CYC. Lenat expresses this as follows:

> 'We can never fully explicate all assumptions underlying the knowledge we express. No matter how hard we try, some assumptions are bound to be left implicit. Contexts, just like the real-world situations they describe, are "rich objects", i.e., there is an infinity of things one could say about a context, and of course only a small finite number of those will ever be explicated.'[190]

Some assumptions are bound to be left implicit and the difficulty that this brings for lifting is unavoidable.

Aside from the issues mentioned above, microtheory technology brought a further challenge. Using lifting, CYC builders were able to create new contexts when they found that the existing contexts did not suite their purposes. Having too many contexts, however, makes it more difficult each time to find the right context for a new piece of knowledge. Furthermore there are costs in logical consistency due to the great deal of lifting that would have been done. At the same time, however, too few contexts give rise to assertions and rules that are huge and difficult to debug. An important challenge

---

[189] D. B. Lenat, DCS, p. 17.

[190] D. B. Lenat, DCS, p. 17.

in incorporating contexts is thus to find a good intermediate size for individual contexts.[191] This challenge has proven difficult to resolve and has greatly influenced the new way that Lenat has adopted to think about contexts. Lenat's new way of thinking is discussed below in the following section.

To sum up, the original strategy that Lenat and his colleagues adopted viewed a context as made up of a series of assertions with a group of associated assumptions. This way of describing contexts was implemented in CYC using a technology developed by Guha to capture knowledge and perform inference using complex data structures called microtheories. This technology included programming for lifting, a means to create and modify contexts by copying and editing assertions and assumptions from one context to another. While microtheory technology facilitated the carving up of the CYC database, important shortcomings arose which undermined and offset its usefulness. Lifting turned out to be a nagging source of logical inconsistencies. Furthermore, as more and more contexts were created it became difficult and tiresome for the knowledge enterer to select the most adequate context each time he/she updated the KB with new information. Moreover, careless knowledge entering was frequently the cause of run time errors which had to be debugged. Finally, Lenat and his team were faced with the difficult challenge of maintaining a good intermediate granularity for the number and choice of contexts.

Due to these shortcomings, the cost of continuing with the original strategy became prohibitively high. Lenat and his crew decided to rethink the way they describe and implement contexts in CYC. Their experience has invalidated their original strategy. A new strategy is underway which is described at length in the following section.

3.2. Rethinking context

The shortcomings in the original strategy motivated Lenat and his team to seek new ways to represent context in CYC. Drawing on the experience gained, Lenat made available to the public an in-depth study of the problems encountered and ways to advance. Lenat proposed an alternative way to represent contexts, and to incorporate them in CYC. Lenat's new proposal is discussed in this section.

3.2.1. The elements of shared context

The surprising reality of contexts, is that while they are complex realities, difficult to understand, our use of them makes our communication less complicated and more expressive. Communication would be difficult and tiresome, if not impossible, if one were obliged to continually explain the context of the phrases that one employs. A rich shared context permits a more economical use of language in which a few phrases convey a great deal of meaning. It permits "terse" signals. To illustrate this point, Lenat uses the case where someone writes or utters a sentence like "Fred told the waiter he

[191] Cfr. D. B. Lenat, DCS, p. 14.

wanted some chips". Lenat observes:

'The reader or listener would be expected to infer many things. To cite just a few of them:

- Fred wants potato chips, not wood chips, cow chips, etc.
- There's no particular set of chips that he wants
- Fred and the waiter were a few feet apart at the time
- This telling event took place in a restaurant
- Fred was a customer dining there at that time
- The waiter was at work there, waiting on Fred at that time
- Fred will start eating the chips very shortly after he gets them
- Fred wants and expects the chips in the next few minutes
- Fred wants and expects the waiter to bring him the chips
- Fred wants and expects a single portion (1-5 oz, 5-25 chips)
- Fred and the waiter speak the same language
- Fred accomplished this by speaking word to the waiter
- Fred might have said "I want chips", or "Chips", or even "Okay"
- Fred assumes the waiter knows/infers all the above things as well
- Fred and the waiter are both human beings
- Fred is old enough to talk (2+ years of age)
- The waiter is old enough to work (4+ years, probably 15+)
- This took place after the date of invention of potato chips (1853)
- "he" means Fred. It is Fred who wants the chips, not the waiter

Etc.'[192]

Humans interact in the real world using terse, syntactically ambiguous phrases such as that in the example above. This is possible because people rely on rich shared contexts to transmit meaning. To further illustrate how shared contexts permit terse signals, Lenat goes on to observe that in response to a question from the waiter, Fred may answer adequately by simply saying "Okay".

Having studied many examples such as the case above, Lenat and his colleagues identify seven elements of shared contexts. In the first place the content of the previous sentences that have just gone by in the dialogue is an important element of shared contexts. In the example above, Fred is able to respond to a subsequent question asked by the waiter with a simple "Okay" because the content of the previous dialogue in

[192] D. B. Lenat, DCS, p. 6.

which Fred placed his order is an element of the shared context when Fred responds to the waiter. Another element of shared context is the form of the previous sentences. The form is the manner in which the signal is formulated. In includes, for example, the choice of words, the sentence structure, the tone and so on.

A third component of shared contexts is the underlying substrate of general knowledge about the real world that we assume that mostly everyone knows. For the purposes of his project, Lenat chose to interpret "general knowledge" as the general knowledge that a modern westerner is assumed to possess. It encompasses things such as recent history and current affairs, everyday physics, household chemistry, famous books, movies, songs and ads, famous people, nutrition, addition, weather, and so on.[193] A large portion of the work done in CYC to better understand contexts has centered around general knowledge.

The fourth element of shared contexts are the simple common sense rules that we rely on for fruitful intercommunication in diverse situations. Lenat explains that they are largely derived from shared experiences such as dining, driving, dating, and so on. In addition to this, common sense rules are often acquired when we learn from misunderstandings, disremembering and similar cognitive failures. Rules are also obtained through shared modes of reasoning. Lenat distinguishes between high modes and lower modes. High modes of reasoning include induction, inspiration and incubation. Lower modes are less rigorous and include modus ponens, dialectic argument, superficial analogy, pigeon-holing etc.[194]

Lenat stresses that his present interests center around general knowledge and the common sense rules that underlie the way we think in ordinary situations. This focus is in keeping with his strategy for CYC. The first goal is to represent and hand-code the general assumptions and rules that we rely on to communicate.[195] These assumptions and rules are stable realities, commonly accepted and taken for granted. Elements of shared contexts that are linked to particular situations in time or to particular conversations are more relevant in the second and third stages of CYC when modules will be added for natural language understanding and learning by discovery.

Though less important for the immediate needs of the first stage of the project, Lenat suggests that the fifth element of a shared context is the current short-term real-world situation, problem, task or environment in which the dialogue takes place, and about which the conversation revolves. An important part of the fifth element are the respective roles that each person participating in the dialogue plays and the assumptions they make about the short term goals and knowledge of the other participants. In contrast to the short term nature of the fifth element, the sixth element of a shared context is the long term background, achievements and roles of the participants in the

---

[193] Cfr. D. B. Lenat, DCS, p. 8.

[194] Cfr. D. B. Lenat, DCS, p. 8.

[195] Cfr. Section 2.5.1.

dialogue which the others are aware of or at least believe to be true. Finally, the seventh element of a shared context is the history of the memorable experiences that the participants shared together. It includes the roles that each played in those events.

As previously mentioned, Lenat's chief interest at present is to better understand and model the third and fourth elements of a shared context so as to advance in the first stage of the project. As fruit of his study and experimentation with many examples, Lenat observes that what mostly matters for these elements is the universality of agreement and not the truth or accuracy of the information that configures common sense reasoning.[196] The assertions "the USA is a big country" and "you should carry a glass of water open-end-up" for example appear imprecise. They are included nevertheless because they enjoy universality of the agreement.

3.2.2. Dimensionalizing context space

In the original strategy to store context, assertions were placed in the most adequate context. If no such context was found a new context was created using lifting. Drawing on his experience Lenat declares this a bad idea. Finding the right context among thousands of possibilities is difficult and time consuming. In addition, as almost every assertion could have its own context, knowledge enterers tended to create many new contexts when updating the KB. This slows problem solving because of the difficulty it brings in finding the right context for seeking solutions. There ought to be a more ordered way to assign contexts and to get around the proliferation of highly specialized contexts. Lenat suggests that this can be achieved by considering contexts in a radically new way. More than just sets of assertions with common assumptions, we can consider the context in which a given assertion applies as being a region in some n-dimensional space. Lenat writes:

> 'When we say (ist P C), we mean that P holds in context C. That is, assertions P is true in context C. But what sort of beast is C? We can reify a context, and then give that specific name in place of C. As we discussed above, that would be a bad idea: almost every assertion in the KB could have its own unique context, even though only a minuscule fraction of those contexts actually merit their own names. On the other hand, we can easily articulate several properties that separate one context from another: level of granularity, time, place, topic, etc. These can be thought of as the labels on different coordinate axes in *n* dimensions. Then each context is a region of that n-space.'[197]

Lenat does not discard completely the previous model in which contexts are considered as nodes in an ontology. He suggests that an alternate strategy is needed to capture aspects of how we use contexts in common sense that allows a more structured, ordered and efficient use. Lenat considers that the best approach is to identify the properties that contexts have in common. These properties are present in each context in

---

[196] Cfr. D. B. Lenat, DCS, p. 9.

[197] D. B. Lenat, DCS, p. 20.

different degrees. If the common properties are considered as coordinate axes in *n* dimensions, a given context corresponds to a region of that n-dimensional space. Lenat summarizes this idea in what he determines to be the first major point of his paper on contexts. He writes:

> 'Point 1: Besides thinking of a context as a "named node in an ontology", let's also think of it as being "a region in some n-dimensional space".'[198]

Regarded merely as a named node, a CYC context is similar to a term in the CYC ontology. One context is related to another by a vague *genlMt* predicate which is used to express the condition of one context being a more restricted set or a more general set with respect to another. This manner of modeling contexts has proven insufficient. In few occasions was one context exactly a more restricted set with respect to another. More often than not, a context was more general or more restricted with respect to another only in a partial way. With respect to some particular properties there may be subset/superset relationships. At the same time, however, with respect to other properties there may be no such relationship at all. The vague *genlMt* predicate thus had many different interpretations according to the contexts it related. Lenat writes:

> 'Thus in one situation genlMt might hold between *The1920sUSA* and *The-TwentiethCenturyUSA* - in this case meaning "is a temporal slice of"; in another situation it might hold between *HumanActivitiesMicrotheory* and *Animal-ActivitiesMicrotheory*, meaning "has a more restricted set of default actors than"; in another situation it might hold between *NewtonianPhysicsMt* and *NaivePhysicsMt*, meaning "is a more precise and accurate theory than"; and so on. In many ways, our naiveté in the early 1990's in having just this one vague *genlMt* predicate is akin to the error that other ontologists make when they have just one *a-kind-of* relation, instead of distinguishing - as we have done for 15 years - the relation *isa (element-of), genls (subset-of), genlPreds (restriction-of-predicate) partOf, subregionOf...*'[199]

The original strategy did not cater well for partial relationships. Knowledge enterers could seldom find contexts that were perfect fits for the assertions they wished to enter. The result was that new contexts were continually created. A great deal of lifting was performed which stalled the updating and inference process later on.

Lenat suggests that the original notion of context in CYC ought to be enriched to better represent partial relationships. These relationships occur in *n* dimensions where *n* is a positive integer denoting the number of largely independent properties that can be used to distinguish one context from another. Considered in this way, the *genlMt* predicate can be affirmed in a partial manner and it is just one of the many relationships that can be affirmed between contexts.

Lenat credits the idea of representing contexts in *n* dimensions to a former Stanford Ph.D. student of his, Paul Cohen. In the mid-eighties, Cohen was a professor at the

---

[198] D. B. Lenat, DCS, p. 20.

[199] D. B. Lenat, DCS, p. 20.

University of Massachusetts. He proposed that the many thousands of relations that we observe and name can be considered as allowed points in an n-dimensional matrix. The dimensions of that matrix are attributes that are true for some predicates and false for others. Lenat recognizes that there is no true limit to the number *n* of dimensions of context space that one could identify. The Empirical Inquiry Hypothesis, however, demands that it ought to be possible to embody and test new models in programs.[200] It is wholly impractical to implement a model for contexts in which contexts have unlimited dimensions. There is a need to choose which raises the question: which dimensions are worth modeling and which are not?

Lenat observes, that not many dimensions are required in order to cater for millions of points in slot space. A dozen or so dimensions, he claims, is more than enough to differentiate the several thousand of slots and predicates that are named in English and in CYC. In addition to this, his years of experience and research has shown that there are high level dimensions that are global points of reference in common sense thought. In the light of this, Lenat proposes that there is a small set of useful, important categories along which context-space can be crudely partitioned.[201] Based on their experience and their principles, Lenat and his team devised a few simple criteria for selecting out this small set. Lenat writes:

> 'The following sort of utility considerations must ultimately be the criteria for deciding which dimensions we support or don't support:
>
> The utility of a dimension should be measured by:
>
> • success in separating out mutually-irrelevant portions of knowledge from each other
>
> • ease of computing the overlap/disjointedness of subsections of the dimension
>
> • how much they tease apart notorious large KB clumps (e.g., HumanActivitiesMt)
>
> • whether the points and regions (especially the extreme ones) correspond to semantically familiar and important real-world concepts and considerations
>
> The utility of a dimension should not be measured by:
>
> • how closely it satisfies some arbitrary philosophical, linguistic, or other model
>
> • whether there happens to be an existing English word succinctly capturing it
>
> • notions of "cleverness", "cleanness", "novelty", "symmetry", "academic

[200] Cfr. section 2.2.1.

[201] Cfr. D. B. Lenat, DCS, p. 23.

interest”

• the need to have 10 dimensions, or 42, or 666, or any particular favorite number.[202]

Using these criteria, Lenat and his team selected out a dozen dimensions for implementation. Lenat explains that they arrived at the figure of a dozen dimensions by beginning with a much larger set of near 100 dimensions. They proceeded to group together pairs of dimensions that were most closely related. The grouping procedure was repeated several times until there ceased to be closely related dimensions. A manageable group of 12 dimensions was chosen.[203]

3.3. The top 12 dimensions of context-space

Lenat lists the 12 groupings that constitute the dimensions of context-space as follows:

‘1. **Absolute time**: a particular time interval in which events occur

2. **TypeOfTime**: a non-absolute type of time period, such as “just after eating”

3. **Absolute Place**: a particular location where events occur, such as “Paris”

4. **TypeOfPlace**: a non-absolute type of place, such as “in bed”

5. **Culture**: linguistic, religious, ethnic, age-group, wealth, etc. of typical actors

6. **Sophistication/Security**: who already knows this, who could learn it, etc.

7. **Topic/Usage**: drilling down into aspects and applications - not subsets

8. **Granularity**: phenomena and detail which are (and are not) ignored

9. **Modality/Disposition/Epistemology**: who wants/believes this content to be true?

10. **Argument-Preference**: local rules for how to resolve pro-con argument disputes

11. **Justification**: are things in this context generally proven, observed, on faith...

12. **Let’s**: local bindings of variables etc. that hold true in that context.’[204]

---

[202] D. B. Lenat, DCS, p. 22.

[203] Cfr. D. B. Lenat, DCS, p. 23.

[204] D. B. Lenat, DCS, p. 24.

In the original strategy, finding the ideal context for a new assertion was difficult and time consuming. In the new model, contexts are mostly specified by just stating or choosing 12 meta-level values for each of the dimensions listed above. Having 12 well defined dimension also speeds up problem solving because search can be localized by favoring the same context in which the problem was framed or some nearby context. Lenat summarizes these advantages in two key points as follows:

> 'Point 2: Placing each new assertion P into the ideal context is expensive today. Idea to speed it up: have it be mostly just stating/choosing 12 meta-level values.
>
> Point 3: Automated inference over a million-rule KB is expensive today. Idea to speed it up: localize the search, favoring the content in same/nearby contexts.'[205]

Lenat admits that a context cannot always be fully specified by assigning values in 12 dimensions. The advantage, however, is that once the values for all 12 dimensions have been specified, then the number of contexts that are further worth distinguishing is dramatically reduced. To illustrate this advantage, Lenat explains that the 12-space region that is defined for an assertion can be considered a context "bucket". The content of the bucket can then be organized into related sub-contexts. He highlights the point that though 12 dimension may appear to be an oversimplification at first, it ought to be taken into account that distinguishing only ten points in each dimension means that trillions of distinct context-spaces can be defined. He insists that if the values in the top 12 dimensions are carefully selected, then other less relevant properties are seldom important for updating the KB and solving problems. The new strategy is to specify the domain assumptions of a context by its 12-space location and where further specification is useful, other domain assumptions can be explicitly specified as CYC formulas. The formulas are listed as *domainAssumptions* of the context.[206]

---

[205] D. B. Lenat, DCS, p. 24.

[206] Lenat identifies eight important phenomenon which have to be handled but which are not worth having as full fledged dimensions. He lists them as follows:

'1. a tag about the probability/certainty/monotonicity of a set of assertions

2. a tag about whether a set of assertions is meant to be run forward or backward

3. a tag about whether typical propositions in some context - which has some time/place/sophistication... interval specified, such as "The1920s" - are claimed to hold throughout that entire interval or just at some points/subsets of that interval. (Or even worse, having a "dimension" like this for each of the 12 other dimensions.)

3.3.1. Dimensions for time and space

The dimensions that occupy the first positions in the list configure time and space. Time and space are two of the most universal notions in CYC. Lenat considers them to be basic notions that we fall back on to intelligently solve problems. In this section, the four dimensions of context-space that configure time and space are described.

1. Dimension: Time

In the new strategy to incorporate context, each context has a value for Time. The assertions that hold true in that context are presumed to occur during that piece of time. Lenat justifies having time as a full fledged dimension by the fact that, in the real world, mostly all that occurs is associated with some time period. He explains, however, that contexts and assertions can be related to time in several ways and that there are difficult issues involved. A simple piece of time is the case of a point or a solid interval. A more complex case is when the pieces are discontinuous such as "every second Wednesday morning". A piece of time can be named or not and can have meaning that goes beyond the mere designation of a time period. (e.g. WorldWarII).[207]

Describing the time dimension, Lenat addresses several difficult issues. A case that lends to confusion is when, for some assertion A1 in some context C with time value t, it is not clear whether it should be entered into the KB as "always-true-during" time *t* or

---

4. other tags, similar to the last one, but talking about whether intervals should be assumed to be open (not including their endpoints) or closed, solid or disjointed, finite or infinite or semi-infinite (infinite in one direction only), and so on.

5. a specific region of any (valid) dimension

6. quoting. This means that we want to represent the fact that $Lenat is a CYC term, and that $Lenat is a person, and yet people are tangible and CYC terms are not. What's the solution? We can use a kind of *quoting* operator; so if we write the expression (quote $Lenat) or '$Lenat, that *represents* the CYC term $Lenat (which them in turn represents the person). Without the quote mark, $Lenat *is* a CYC term and represents the person, not the CYC term.

7. Physical attributes, such as: Temperature, Pressure, Orientation, Stability, etc. of the typical objects and events talked about in this context.

8. The level of "anthropacity" of the context, by which we mean: what does it presume that there "is" in the world: a physical universe? Like ours? People? Human history as we know it? Etc. This can largely be computed from the various assertions (and assumptions) of the context. E.g., if assertions talk about I Love Lucy shows, then the world must be pretty close to ours.' (D. B. Lenat, DCS, p. 25).

[207] Cfr. D. B. Lenat, DCS, p. 27.

"sometimes-true-during" time t. To illustrate this difficulty, Lenat uses as an example the assertion: "Jane was married during the 1980's". It is not clear whether the assertions is always true during the time period 1980-1990 or only for some time during the 1980's. The difficulty becomes more complex with the need to also distinguish "true-in-time-x" from "relevant-in-time-x". Lenat illustrates this with the example sentence: "Bill Clinton is the President of the USA during 1997." The statement is true for all time. It is only relevant, however, during and after 1997. Before then no one acted differently because of this fact and it was therefore irrelevant.[208] Based on his experience building CYC, Lenat judges the true-in-time versus relevant-in-time distinction to be sufficiently important to warrant two time dimensions, one for truth and one for relevance. Taking relevance into account enables the search space for browsing and inferencing to be radically reduced, hence its importance. Putting relevance and duration together, Lenat explains that in the new approach, there are four discernible regions in each context from the time value perspective. These regions correspond to four related functions in CYC. The following diagram presents these functions and illustrates the genlMt relationships between the four regions of a context C. Each of the microtheories contain a subset of the assertions of the microtheory above it.

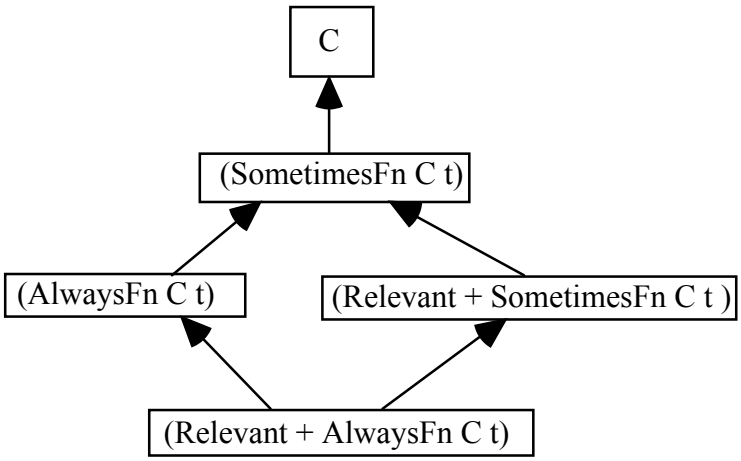


Figure 3. The four regions of context in the time dimension

The four regions of context from the time perspective means that defining a time value for a context is equivalent to specifying four contexts. As a simple example, to have a context for the 1920's in the USA it is as if there were four contexts:

RelevantThroughout1920sUSAMt,

RelevantSometimeIn1920sUSAMt,

TrueThroughout1920sUSAMt and

TrueSometimeIn1920sUSAMt.

---

[208] D. B. Lenat, DCS, p. 30.

Lenat stresses, however, that it will be very unusual for knowledge enterers to hand-assert anything in the latter two contexts. This is so because if an assertion is not relevant then it is normally not worth hand asserting it into that context.[209]

Together with the regions, there ought to be a notion in each context of what the meaningful unit of time is understood to be. Lenat terms this the "temporal granule size". It indicates how small a piece of time is before it is considered an indivisible moment. To illustrate this point, Lenat observes that the granule of time for a macroeconomics theory of nineteenth century Europe might be a day or even a year. For a track and field event, it might be 0.01 second. Granules of time often map to some absolute calendar-type time and a single context can have granules of different sizes.[210]

Another reality that ought to be carefully discerned is temporal projection. The temporal projection of an assertion A1 that holds at time *t* is the period of time after *t* that it is likely to continue to hold and the period of time before *t* that it is likely to have already been true. To simplify data capture, CYC ought to be furnished with a means for specifying, determining or calculating the temporal projections of assertions. Lenat suggests that temporal projection can be adequately represented using simple probability or likelihood distributions over time, both before and after the stated time period in which an assertions is said to hold. He declares this the fourth key point of his new strategy as follows:

> 'Point 4: Each assertion points to one of a handful of persistence distributions (spike, step, uniform, normal, etc.) and gives a crude estimate of the parameters of that distribution (e.g., mean and standard deviation in the case of a normal distribution, or max./min. and overall length in the case of a

---

[209] Cfr. D. B. Lenat, DCS, pp. 32-33.

In simple prose Lenat describes the four functions as follows:

'($Relevant+AlwaysFn C t) = C1 means that the assertions in context C1 are precisely those assertions that hold in C and which happen to be both true and relevant throughout the entire temporal extent of the Temporal Thing t.

($Relevant+SometimesFn C t) = C2 means that the assertions in context C2 are precisely those assertions that hold in C and which happen to be both true and relevant during at least *some* portion of the temporal extent of t (a $TemporalThing).

($AlwaysFn C t) = C3 means that the assertions in context C3 are precisely those assertions that hold in C and which hold true throughout the entire temporal extent of t.

($SometimesFn C t) = C4 means that the assertions in context C4 are precisely those assertions that hold in C and which are true during at least *some* portion of the temporal extent of t.' (D. B. Lenat, DCS, p. 32).

[210] Cfr. D. B. Lenat, DCS, p. 30.

less regular one).'[211]

When a context is defined, default distributions are specified for each of the four time regions. This facilitates the process of defining the temporal projections for assertions in that context later on.

Another problematic issue is the fact that the time information for assertions is often relative, not absolute. The time information is relative when the location in time is specified relative to the duration of other events (e.g. "the time it takes to walk to the library") or relative to the starting and ending time points (e.g. "when he arrives/leaves").

To efficiently solve problems in a manner that is not trivial, the new strategy that Lenat proposes needs ways to cater for the several types of meta-level time information that assertions can contain. How does CYC decide whether assertion A1 is true sometimes-during or always-during the time period of context C1? How does CYC decide whether it applies to some particular piece of time, and/or standing in some relation to some other event or piece of time? How does CYC decide how long into the future that the assertion will remain true? Lenat explains that it will almost always be possible to infer such information from the predicates used, the other terms that are employed in the assertion and the default values attached to the context as a whole. The result of this is that only rarely will the ontological engineer have to explicitly specify the parameters for meta-level information about assertions. This is an important claim which Lenat presents as the fifth main point of his new proposal. He writes:

> 'Point 5: Most assertions' persistence-distributions are "inherited" from a single default distribution attached to their context (and, in turn, even that is typically inherited/inferred); a few assertions get this overridden by dint of their particular predicates (such as gender and birthDate) [usually their main predicate] possibly with some typing of arguments to that predicate. And, finally, there is the very tiny fraction of the KB's assertions that need to have a manually-entered persistence-distribution. A similar tripartite scheme (context-wide unless predicates-involved unless idiosyncratic-for-that-assertion) applies to deciding if it is sometimes-true-during C1 versus always-true-during C1; etc.'[212]

Point five refers to temporal projection and the sometimes-true-during vs. always-true-during meta level values. Lenat proposes that the tripartite scheme also applies to the absolute and relative specification of the time interval. He claims that in most cases, the system ought to be able to guess both the duration and starting/ending times based on the terms used in the assertion. As a simple illustration, Lenat writes: 'In the case of "George Washington wears wooden teeth", e.g., the time period of the objective content is during that person's lifetime, and also during the time period of commercial use of

---

[211] D. B. Lenat, DCS, p. 28.

[212] D. B. Lenat, DCS, p. 29.

wooden teeth.'[213] Going a step further, he claims that a similar tripartite strategy applies, not just for the time dimension, but to most aspects of most of the dimensions of context-space.[214]

2. Dimension: TypeOfTime

The principal elements of the upper CYC ontology were discussed in the second chapter. In this scheme $TemporalStuffType and $TemporalObjectType were considered important subsets of $StuffType and $ObjectType respectively. $Temporal-StuffType is the collection of temporal realities that are stuff-like and $TemporalObject-Type is the collection of temporal realities that are object-like as opposed to stuff-like.[215] The collection $TemporalThing gathers together the collection of all individual things which have a particular temporal extent. It is an element of the collection $TemporalStuffType which means that it can be understood to be a collection of individual pieces of temporal stuff. The dimension TypeOfTime of a context, specifies the default type of $TemporalThing in which the actions take place, the actors are situated, etc. for the assertions in that context. Lenat explains that they usually turn out to be instances of $TemporalObjectType though values along this dimension are at times non-atomic terms (NATs). He writes: 'Other possible values along this dimension will be NATs (non-atomic terms) constructed by applying some sort of "during" operator to some event or process: for example: during an intermission, during an elections campaign, and so forth.'[216] Some legal values for TypeOfTime which are $TemporalThings and also $TemporalObjectType are, for example: $1998Winter-OlympicEvent, $USCabinetMeeting and $ChristmasHoliday[217]

---

[213] D. B. Lenat, DCS, p. 29.

[214] Cfr. D. B. Lenat, DCS, p. 29.

[215] Cfr. Section 2.4.3.1. It was explained that something is temporally stuff-like, when it has the property that when a member of the class is temporally carved into pieces, each one is still an instance of the original class (e.g. The class $AnimalWalkingProcess). Something is temporally object-like when it has the property of being an individual temporal reality. The characteristic property of stuff-like things does not apply (e.g. $WalkingToTheMailboxAndBack). An object typically inherits its intrinsic properties from the type of substance (stuff) of which it is an instance, and it inherits extrinsic properties from the type of object that it is an instance of.

[216] D. B. Lenat, DCS, p. 37.

[217] Cfr. D. B. Lenat, DCS, p. 37. Lenat also provides a list of several subsets of $TemporalObjectType, instances of which could be legal values for TypeOfTime. The subsets named are as follows: $SportsCategoryType, $AnnualEventType, $WeeklyEventType, $ClimateCycleType, $Exclusive-Treatment, $CourseOfStudyType, $DrugAdministrationRouteType, $Cal-endarCoveringType, $DayOfTheMonthType, $TimeOfDayType,

The dimension TypeOfTime is relatively independent of Time. Values along the dimension TypeOfTime are pieces of temporal stuff that are usually $TemporalObject-Type. They associate the assertions, actors, etc. in a given context with elements in the CYC ontology. These elements are instances of collections which in turn are subsets of other collections. This establishes a partial order between contexts in the TypeOfTime dimension. As in the case of Time, to cater for relevance and duration, four functions have been defined in CYC to capture TypeOfTime values as follows:

($Relevant+AlwaysFn C TypeOfTime) = C1

($Relevant+SometimesFn C TypeOfTime) = C2

($AlwaysFn C TypeOfTime) = C3

($SometimesFn C TypeOfTime) = C4.[218]

To illustrate the meaning of C1 and C2, Lenat explains that C1 could be used to express the idea that an assertions P is true during every moment of every intermission or during every instant of every Friday, etc. C2 could be used to express the idea that P is true at some time during an intermission or at some time during a meal, or at some time on Friday, etc.[219]

$TemporallyDisjointedIntervalType, $ConveningEventType, $DayOfYear-Type, $MonthOfYearType, and $CalendarSeasonType.

[218] Lenat describes the functions as follows:

'($Relevant+AlwaysFn C TypeOfTime) = C1 means that the assertions in context C1 are precisely those assertions that hold in C and which happen to be both true and relevant throughout the entire temporal extent of each temporal object t which is an instance of TypeOfTime (which must itself be a $TemporalThingType and is usually a TemporalObjectType)...

($Relevant+SometimesFn C TypeOfTime) = C2 means that the assertions in context C2 are precisely those assertions that hold in C and which happen to be both true and relevant during at least *some* portion of the temporal extent of each t which is an instance of TypeOfTime...

($AlwaysFn C TypeOfTime) = C3 means that the assertions in context C3 are precisely those assertions that hold in C and which hold true throughout the entire temporal extent of every t which is an instance of TypeOfTime...

($SometimesFn C TypeOfTime) = C4 means that the assertions in context C4 are precisely those assertions that hold in C and which are true during at least *some* portion of the temporal extent of each temporal object t which is an instance of TypeOfTime' (D. B. Lenat, DCS, pp. 37-38).

[219] Cfr. D. B. Lenat, DCS, pp. 37-38.

3. Dimension: GeoLocation

Like time, space is one of the most universal notions in CYC. Because of its universality, the spatial characteristics of things are important defining characteristics for distinguishing one thing from another. An important spatial characteristic of things are their geographical location. Lenat places it second on the list of the top 12 dimensions that configure context space. The GeoLocation of a context captures the default geographical location of the actors, events, etc. that are mentioned in the assertions in the context.[220] In his new strategy, Lenat represents the geographical location of a context in a manner that is analogous to the time dimension. In the time dimension, for some assertion A1 in some context C with time value t, it ought to be carefully discerned whether A1 should be entered into the KB as "always-true-during" time *t* or "sometimes-true-during" time t. In an analogous manner, in the GeoLocation dimension, for some assertion A1 in context C with spatial region r, it ought to be carefully discerned whether A1 should be entered in the KB as "true-everywhere" in r, or "true-somewhere" in r.

Together with the always-true/sometimes-true distinction in the time dimension, Lenat stressed the need to distinguish between the portions of the content of a context that are *just* true in time *t* or both true and relevant in time t. In an analogous manner, in the GeoLocation dimension, one portion of the content of the context is just true for some spatial region, and the rest is both true and relevant. As in the time dimension, each context is carved into four discernible regions which correspond to four related functions in CYC. The following diagram presents these functions and illustrates the genlMt relationships between the regions. Each of the microtheories contain a subset of the assertions of the microtheory above it.[221]

---

[220] Cfr. D. B. Lenat, DCS, pp. 39.

[221] Lenat describes the four functions in ordinary prose as follows:

'($Relevant+EverywhereFn C t) = C1 means that the assertions in context C1 are precisely those assertions that hold in C and which happen to be both true and relevant throughout the entire spatial extent of the $SpatialThing t.

($Relevant+SomewhereFn C t) = C2 means that the assertions in context C2 are precisely those assertions that hold in C and which happen to be both true and relevant during at least *some* portion of the spatial extent of t (a $SpatialThing).

($EverywhereFn C t) = C3 means that the assertions in context C3 are precisely those assertions that hold in C and which hold true throughout the entire spatial extent of t.

($SomewhereFn C t) = C4 means that the assertions in context C4 are precisely those assertions that hold in C and which are true during at least *some* portion of the spatial extent of t.' (D. B. Lenat, DCS, p. 40).

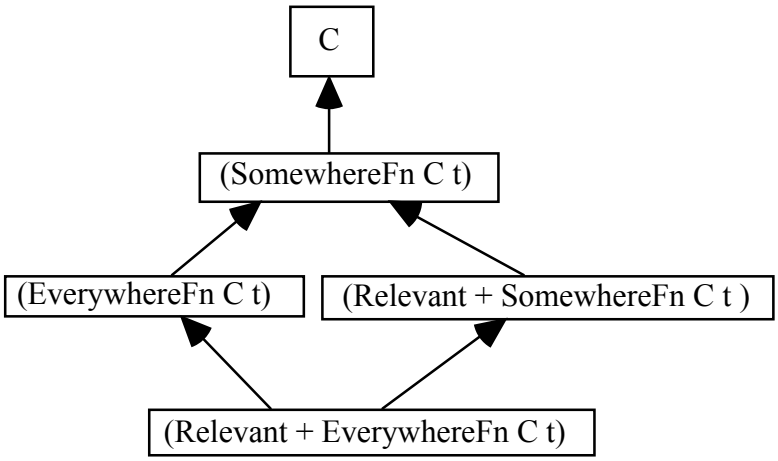


Figure 3. The four regions of context in the GeoLocation dimension

Having four pieces for each GeoLocation *r* means that defining a spatial region for a context is equivalent to specifying four contexts. As a simple example, to have a GeoLocation context for the 1920's in the USA it is as if there were four contexts:

TrueEverywhereIn1920sUSAMt,

RelevantEverywhereIn1920sUSAMt,

TrueSomewhereIn1920sUSAMt and

RelevantSomewhereIn1920sUSAMt.

As with time, together with the four regions, another component of each value along the GeoLocation dimension is the granule size. This indicates how small a piece of space is in that context before it is considered an indivisible point. The piece of space that is associated is the size of some spatial object *t* or, when the right size is not known, a CYC collection of granules can be associated.

Carrying the analogy with time another step further, Lenat indicates that in the new strategy, each value along the GeoLocation dimension will have a persistence distribution. It is a default for how likely it is that a typical assertion that is made in context C will be true somewhere outside the GeoLocation of C. When a given assertions has a persistence distribution that is different from the default distribution of the context in which it is placed, the distribution of that assertion overrides the default.[222]

4. Dimension: TypeOfPlace

Just as GeoLocation was analogous to Time, TypeOfPlace is analogous to TypeOf-Time. To describe the dimension TypeOfPlace, Lenat refers to the CYC collection $SpatialThing. It is the collection of all individual things which have a particular spatial extent. Subsets of $SpatialThing are elements of the collection $SpatialThingType

[222] Cfr. D. B. Lenat, DCS, p. 41.

which means that $SpatialThing can be understood to be a collection of individual pieces of spatial stuff. The dimension TypeOfPlace of a context, specifies the default type of $SpatialThing in which the actions take place, the actors are situated etc. for the assertions in that context. Lenat explains that they usually turn out to be instances of $SpatialObjectType.[223] In a cooking context, for example, the TypeOfPlace value might be Kitchen. In a ModernWesternSleeping context, the TypeOfPlace might be Bedroom and HotelRoom. Lenat notes that in a Japanese culture context, the TypeOfPlace for a sleeping context would not be Bedroom because the Japanese sleep in the living room. This illustrates the point that the cultural setting of a context is an important factor for determining the most adequate values for TypeOfPlace. While this is so, some values for TypeOfPlace are by nature culture independent. Such is the case with values such as Underwater, Underground, Arboreal, Airborne, Outdoors, a dark place, etc.[224]

The values for TypeOfPlace, like TypeOfTime, are instances of collections which in turn are subsets of other collections. This induces a partial order among the values in the TypeOfPlace dimension. Lenat explains that another important way to introduce partial order is by means of the $physicalDecomposition predicate. This predicate can be used to relate values along the TypeOfPlace dimension axes. It expresses the relation that one thing is a physical decomposition of another. As in the case of GeoLocation, to cater for relevance and the everywhere/somewhere distinction, four functions have been defined in CYC to capture TypeOfPlace values as follows:

($Relevant+EverywhereFn C TypeOfPlace) = C1

($Relevant+SomewhereFn C TypeOfPlace) = C2

($EverywhereFn C TypeOfPlace) = C3

($SomewhereFn C TypeOfPlace) = C4.[225]

---

[223] Cfr. D. B. Lenat, DCS, p. 44. In a note on page 41 Lenat clarifies that the terms $SpatialThing, $SpatialThingType and $SpatialObjectType do not exist in the current CYC vocabulary. These terms, he affirms, ought to be added to the ontology or substituted by equivalent terms before the official publication of the document.

[224] Cfr. D. B. Lenat, DCS, p. 44.

[225] Lenat describes the functions as follows:

'($Relevant+EverywhereFn C TypeOfPlace) = C1 means that the assertions in context C1 are precisely those assertions that hold in C and which happen to be both true and relevant throughout the entire spatial extent of each spatial object x which is an instance of TypeOfPlace (which must itself be a $SpatialThingType and is usually a $SpatialObjectType)...

($Relevant+SomewhereFn C TypeOfPlace) = C2 means that the assertions in context C2 are precisely those assertions that hold in C and which happen to be both true and relevant at least somewhere in the temporal extent of every x which is an instance of TypeOfPlace...

The Dimensions Time, TypeOfTime, GeoLocation and TypeOfPlace can be brought together in a powerful manner to configure the context for a group of assertions. Lenat presents the case of a context about "daytime driving automobiles in rural 1920's America". He succinctly describes how dimensions can be used to factor out the elements of the context, and puts forth this description as the sixth important idea of his new strategy to represent context. He presents his point as follows:

> Point 6: Starting with a context about "daytime driving automobiles in rural 1920's America..." we can use the Time dimension to factor out the "1920's" part of what the context was, the TypeOfTime dimension to factor out the "daytime" part, the GeoLocation dimension to factor out the "America" part, and we will use the TypeOfPlace and Topic dimensions, respectively, to filter out the "rural" and "driving automobiles" parts. In this case, the whole context will be uniquely specified by specifying a set of values on a handful of our dimensions. There may not be any extra assumptions at all that have to get stated; the context C may just receive some assertions. In other words, C is just a set of assertions plus a set of values along a few of our dimensions.'[226]

By carefully choosing values along a few dimensions, a context can be almost completely defined. To complete the procedure, for each of the elements that are factored out and for each of their automatically defined subdivisions, default persistence distributions have to be defined. Lenat explains that in most cases an adequate persistence distribution can be specified by selecting from a list of distributions that CYC provides. After selecting the desired distribution, the user ought to modify the required parameters according to the demands of the context in question.[227]

Lenat upholds that proceeding in this way has many advantages over the original strategy. In the first place having a common structure for all contexts makes it easier to specify new contexts and to relate them during inference. Furthermore, the use of functions is an effective way to correct the proliferation of reified contexts that occurred in the original strategy. Where reified contexts would have been used in the original strategy, functional definitions or "settings" can be employed. This permits fewer reified contexts which simplifies keying-in and problem solving later on. Lenat summarizes this benefit as the seventh key point of his proposal as follows:

($EverywhereFn C TypeOfPlace) = C3 means that the assertions in context C3 are precisely those assertions that hold in C and which hold true throughout the entire spatial extent of every x which is an instance of TypeOfPlace...

($SomewhereFn C TypeOfPlace) = C4 means that the assertions in context C4 are precisely those assertions that hold in C and which are true at least some portion of the spatial extent of every spatial thing x which is an instance of TypeOfPlace' (D. B. Lenat, DCS, pp. 44-45).

[226] D. B. Lenat, DCS, p. 42.

[227] Cfr. D. B. Lenat, DCS, pp. 33-35.

'Point 7: Most "settings" for a context - for any one full value of any one single dimension of context-space - are not worth naming or remembering.'[228]

Lenat acknowledges that not all contexts can be adequately specified by assigning values along 12 dimensions. More complicated specifications are at times required to properly capture the properties and characteristics of a context. He explains that the $domainAssumptions predicate can be used in these cases to specify additional constraints. Sometimes it may be necessary to reify a new context but this should be kept to a minimum. Lenat suggests that an artful use of the $domainAssumptions predicate can often eliminate the need to reify new contexts.[229]

3.3.2. Eight other useful dimensions

Together with the four dimensions for space and time, Lenat identifies eight other dimensions which he considers to be most useful for representing contexts and performing inference using contexts in CYC. In this section, these eight other dimension are briefly described.

1. Dimension: Culture

In order to be useful, a context for a set of assertions ought to include information about the cultural background or setting to which the intelligent actors in the assertions belong. Knowing whether the actors are male or female, young or old, atheist or theist, American or Chinese for example, will often affect how problems ought to be resolved when questions are asked to CYC later on. The dimension Culture is meant to specify the default values for such information. To facilitate data entry and inference, Lenat explains that the possible points of the Culture dimension are organized along 12 sub-culture types or "sub-dimensions" which are somewhat independent. The Culture constraint is specified via a predicate called $mtAgentType using an expression of the form: ($mtAgentType C HCT) where C is a context, and HCT is a $HumanCultureType.[230]

---

[228] D. B. Lenat, DCS, p. 47.

[229] Cfr. D. B. Lenat, DCS, p. 48.

[230] Cfr. D. B. Lenat, DCS, p. 47. On the same page, Lenat lists the 12 principle sub-dimensions of culture as follows:

'political culture (Democrat, Fascist, Libertarian,...)

sexual culture (male, female)

sexual orientation culture (heterosexual,...)

age culture (young,..., adult,...,old)

generation culture (Renaissance, of the 1960's, generation X...)

religious culture (Atheist, Fundamentalist Jew, Muslim, Catholic,...)

Using multiple $mtAgentType definitions, contexts that are combinations of various sub-culture values can be easily defined. Multiple $mtAgentType assertions are interpreted as conjoining and the resulting default can be reified if necessary. Sometimes a single context may have two or three specific Cultures whose union (as opposed to intersection in the case of conjoining) is meant to be the Culture of the context. To cater for these cases, a special predicate, ($cultureUnion X Y) has been added to CYC.[231]

2. Dimension: Sophistication / Security

The dimension Sophistication/Security follows Culture. The Sophistication of a context captures the characteristics that an intelligent subject ought to possess in order to know the content of the assertions in that context. To illustrate this idea, Lenat writes:

> 'A detailed medical context might contain information known only by physicians who specialize in that area. A company's salary data might be known only by Human Resources staff of that company. A listing of the latest stock tips might be known only to those who have paid for them.'[232]

Granted that a certain sophistication is required to know the assertions in a context, a strategy is required to represent and store such information. Lenat claims that sophistication can be adequately described by means of a linear scale or by choosing values from a well-defined set of possible settings. The linear scale and the set of attributes that are employed are similar to those used in the Culture dimension. Choosing the appropriate attribute settings for the sophistication value of a context at first appears to be a time consuming and difficult task. Lenat explains, however, that, as with most other dimensions, the process is made easier by relying on copy/edit functions and automatic inference. Most of the time, the user will not need to manually specify values. The settings can be copied from similar or related assertions or they can be automatically inferred by the system based on information about the prior context, the current task, the sophistication of the user who is making assertions and the form and terms of the assertions that are being made.[233]

---

ancestral culture (of Irish descent,...)

geo-political culture (American, European,...)

regional culture (midwestern, east-coast, west coast,...)

region-type culture (urban, rural,...)

economic/work culture (idle rich, working academician, Protestant moralists,..)

legal culture (based on aspects of the codes of conduct/enforcement)'

[231] Cfr. D. B. Lenat, DCS, p. 47.

[232] D. B. Lenat, DCS, p. 48.

[233] Cfr. D. B. Lenat, DCS, p. 49.

Lenat puts the notion of "Security" together with Sophistication. Like sophistication, security refers to characteristics that the users must possess in order to make use of the assertions in a given context. Whereas sophistication captured attributes of persons who know the group of assertions, security refers to the authorization that is required in order to see, use or modify the assertions in a given context. As with sophistication, security is defined by specifying appropriate settings. Using copy/edit functions and automatic inference the process of defining security defaults is greatly simplified. Sophistication and Security are specified in CYC using the same quinary predicate $mtAccessConstraint.[234]

3. Dimension: Topic

For a given context, the dimension Topic answers the question of what it is about. Lenat considers the dimension topic to be particularly useful for separating out large groups of assertions in the CYC KB, such as the content of the current huge microtheory bucket for human activities.[235] As with other dimensions, the Topic for a context is defined by specifying the appropriate settings. Lenat explains that the attributes for Topic is simply a set of CYC terms. This means that all the assertions in a context are about one or more of the terms in the Topic settings. Lenat clarifies that an assertion is about one or more terms when the content of that assertion explicitly mentions those terms or instances of them. He recognizes that as new contexts are specified, new terms may need to be created in order to demarcate topics that are worth distinguishing. This, however is a welcome tradeoff for the benefits that come with well-defined Topic settings.[236]

One of the benefits of having well defined Topic settings is that it makes it easier to relate contexts according to what they are about . Lenat explains that to relate contexts in the Topic dimension, there is a new CYC predicate $subTopic. It is a binary predicate of the form ($subTopic C1 C2) which means that context C2 is a subtopic of context C1. The notion of sub-topic, however, should not be confused with subset. Lenat claims that to determine what something is "about", we normally ask the same type of questions. These questions, Lenat suggests, can be listed and categorized to facilitate the definition of topics and subtopics. Defending this point Lenat writes:

> 'There are a relatively small number of questions which characterize what we colloquially mean by a topic being "about" x... All of those questions can also lead from a topic to one of its subtopics.'[237]

---

[234] Cfr. D. B. Lenat, DCS, pp. 49-50.

[235] Cfr. D. B. Lenat, DCS, p. 51.

[236] Cfr. D. B. Lenat, DCS, p. 52. In CYC a context is related to a term using the new predicate ($propositionalInfoAbout C TERM) which means that the context C is about the reifiable term TERM.

[237] D. B. Lenat, DCS, p. 52.

Lenat goes on to provide a long list of examples of such questions distinguishing between short-term practical tasks at hand, such as opening a door and relatively long-term tasks at hand. He also provides an additional list of questions for cases where the topic is a task or a tangible thing in the world.[238]

4. Dimension: Granularity

The dimension Granularity follows Topic. The notion of Granularity captures how small the pieces of the objects, events, time-periods etc. are before they are considered indivisible. Lenat explains that though one can generally structure things into a hierarchy of integrative levels (for example moving from molecular to cellular to organ to organism to society level for living systems), assertions at one level do not simply translate up or down to the next level. Each level is almost its own world, and this, Lenat affirms, is the reason for having the Granularity dimension.[239]

Lenat proposes that Granularity can be adequately specified using six binary predicates. These predicates can be understood as sub-dimensions for specifying the Granule value of a context. The first three predicates define spatial granules. The first defines boundaries for distance outside of which distance values are to be considered irrelevant. The second specifies limits for volume. Objects with volumes outside these

---

[238] Some of the questions that Lenat proposes in the three areas are as follows:

1. If the topic is some relatively short-term practical task at hand:
inputs needed (including energy, information, and physical materials)
tools and equipment required/involved
physical situation, configuration, and location
currently governing codes and standards
persons/interests to be served
financial/commercial consequences of the task
...
2. If the topic is some relatively long-term task at hand, then include:
development of methodology over time
commercial practice in the field
spectrum of participants and players
changes in governing codes and standards
...
3. If the topic is a task or a tangible thing in the world then include:
the origins of the task/thing
how it was dealt with historically
literary or cultural role of the task type or thing type
implications for philosophy, science, history
are the ultimate goals of the task/thing really worth pursuing?
... (D. B. Lenat, DCS, p. 52).

[239] Cfr. D. B. Lenat, DCS, p. 54.

limits have no bearing in that context. The third defines the range for area that is to be considered pertinent. Together with the spatial predicates, three other predicates configure Granularity in CYC. The first concerns the level of abstraction that should be taken into account when reasoning in a context. Level of abstraction captures the specificity of classes (collections), relationships, measurements, etc. that are relevant for a context. Lenat explains:

> 'The context may assume that the focal actors are at a certain level such as people, corporations, countries; a book versus an edition of a book versus the content of the book; a proposition or an entire theory or a representation of one of those two in some language; particular versus abstract game pieces; etc.'[240]

The notion of level of abstraction that the example illustrates is represented in CYC by a term in a hierarchy of subset/superset relationships. This hierarchy is already present in CYC as a result of the $genls relations that exist between collections. To specify level of abstraction settings the binary predicate $mtSubsetAbstractionLevel is used.[241]

Lenat explains that, in CYC, another hierarchical structure can be considered to be present due to the $parts relationship between collections. The $parts predicate structures CYC terms into a hierarchy according to their size. Lenat suggests that another sub-dimension of Granularity is the level of abstraction according to the parts hierarchy. This sub-dimensions indicates the size of objects, events, object-parts, time-periods, subevents, organizations etc. that are relevant in the context under consideration. The binary predicate that captures this notion in CYC is $mtPartonomic-AbstractionLevel.[242]

The final sub-dimension of Granularity captures the "phenomenon" that are relevant in a context. Lenat explains that this dimension arises by asking the question "what phenomenon matter?" The phenomenon that are pertinent help to discern the best size for the granules and the types of granules that should be taken into account. Phenomenon is specified using the binary predicate ($mtRelevantPhenomenon C

---

[240] D. B. Lenat, DCS, p. 54.

[241] In more technical terms Lenat states and describes the level of abstraction predicate as follows:

'($mtSubsetAbstractionLevel C Col) means that when reasoning in or with the context C, (instances of) classes of objects at about the level of abstraction indicated by the collection Col. should be attended to, or taken into account. If you consider the knowledge base as a lattice of abstract concepts arranged in a hierarchical structure via the subset/superset relation ($genls), then the further away (up or down) in the hierarchy you go from Col, the less likely it is that (instances of) those classes should be used in inferencing in context C.' (D. B. Lenat, DCS, p. 55).

[242] Cfr. D. B. Lenat, DCS, p. 56.

TERM). It means that when reasoning in or with the context C, the phenomenon indicated by the reifiable term TERM should be attended to, or taken into account.[243]

5. Dimension: Modality / Disposition / Epistemology

Modality, Disposition and Epistemology are grouped together in the new strategy to form the ninth dimension of context space. Lenat explains that these notions were grouped together because they are closely related. The first notion, Modality, refers to the default modal status of assertions that are made in the context. Modality, Lenat explains, answers the interrogative: "Are they beliefs, agreements, expectations, memories, etc., and if so, whose?"[244] The second notion, Disposition, refers to the default attitudes or reactions of the actors in the assertions to the case that the assertions are true. Lenat explains:

> 'Suppose an agent believes assertion P to be true; that's modality. Now suppose they are *disposed* to think that it is a good thing or a bad thing, that is disposition. The disposition of assertion P refers to the attitude of agents (who believe, expect, etc. P) toward P being true. Thus coupled with modality, disposition includes dreads, desires, goals, annoyances, etc.'[245]

Disposition concerns attitudes and is thus particularly useful for reasoning about behavior, moods and motivations.

The third notion, Epistemology, captures the sense in which something is meant to be understood or in which it can be interpreted. Lenat describes it as follow:

> 'Epistemology refers to the default epistemological status of assertions locally made in this context: Are they intended/taken to be true, intentional lies/works of fiction/humor/erotica, rumors etc. Or, alternatively: who believes these things to be true? Who believes them to be fictional? To be a joke? and so on.'[246]

Epistemology is important for reasoning about lies, jokes and misunderstandings. Regarding their technical definition in CYC, Lenat explains that Epistemology, Modality and Disposition share the same collections and predicates.[247]

---

[243] Cfr. D. B. Lenat, DCS, p. 56.

[244] D. B. Lenat, DCS, p. 57.

[245] D. B. Lenat, DCS, p. 58.

[246] D. B. Lenat, DCS, p. 58.

[247] In simple prose Lenat describes them as follows:

'$PropositionalAttitudeContext - The collection of all contexts which contain assertions expressing the propositional attitudes of some agent...

$BeliefsContext - The collection of all contexts which contain assertions expressing the beliefs of some agent or group of agents...

6. Dimension: Argument-Preference

The Argument preference for a context is the default set of heuristics that are used in the context to resolve pro/con arguments. The heuristics that are specified have weights to indicate the order in which they should be brought to bear to resolve arguments. Lenat explains: 'suppose you have a short novice argument and a long expert argument: do you prefer "prefer short arguments to long ones" to "prefer expert arguments to novice ones"?'[248] In CYC, the settings for Argument-Preference are instances of the new collection called $ArgumentPreferenceSet.[249]

7. Dimension: Justification

The Justification for a context, writes Lenat is information about the 'justification, kinds of justifications, sets of justification, or sets of kinds of justifications that justify all/most/some of the assertions in that context.'[250] This dimension makes it possible to group together assertions that are justified in a similar manner into one context. When a new assertion is added to that context, CYC automatically knows what types of

---

($believesMt AGT C) means that the Agent AGT's beliefs include all of the propositions in context C... In general C will be a BeliefsContext...

$DesiresContext - The collection of all contexts which contain assertions expressing the desires of some agent or group of agents...

($desiresMt AGT C) means that the Agent AGT's desires include all of the propositions in the context C...In general C will be a $DesiresContext.

$GoalsContext - The collection of all contexts which contain assertions expressing the goals of some agent or group of agents...There are two major types of goals: those to sustain something that is already true, and those to attain something desirable but currently not true...

($goalsMt AGT C) means that the Agent AGT's goals include all of the propositions in context C.' (D. B. Lenat, DCS, p. 59).

[248] D. B. Lenat, DCS, p. 61.

[249] Cfr. D. B. Lenat, DCS, p. 61. Lenat explains that instances of this set are partially ordered sets of heuristics for different contexts. They contain assertions of the two following forms:

P => (prefer A1 A2) where P is some condition involving arguments A1 and A2. E.g., (shorter A1 A2) => (prefer A1 A2)

P1 => (overrides Pref1 Pref2) where Pref1 and Pref2 are each statements of that previous form (such as (shorter A1 A2) => (prefer A1 A2).

Assertions of type P indicate which arguments are to be preferred when a pair of arguments A1 and A2 clash. Assertions of type P1 specify the order in which these criteria ought to be applied.

[250] D. B. Lenat, DCS, p. 62.

justification that assertion has. An assertion may be justified, for example, by its cause, by statistical evidence, by intuition, by faith, by assumption, etc. This information is useful for solving problems about why something is held to be true. In addition to this, having Justification defaults for groups of assertions makes it easier to perform truth maintenance when assertions that underlie or justify others are retracted.[251]

8. Dimension: Let's (and misc. domain assumptions)

The last of the twelve dimensions gathers together miscellaneous assertions that are assumed to hold true in a given context. Lenat calls this dimension the "Let's" dimension because in ordinary prose assumptions are often formulated using the syntactic form "Let...". Lenat offers two examples: 'Let X be a person and suppose Y is the person's height' or 'Let Z be a right triangle, whose hypotenuse is of length h.'[252] Lenat explains that this dimension is important because it pre-defines a set of variables and formal relations among them. The assertions in that context can refer to the pre-defined variables and CYC automatically deduces many other new relationships that ought to be set up. This facilitates the task of updating the KB and performing inference later on. Assumptions in the Lets dimension are defined using the *$domainAssumptions* predicate. Lenat illustrates how this predicate can be used to configure the Let's dimension of a context as follows:

> '(domainAssumptions C P)- This means that P holds in C, by assumption...We could assert ($domainAssumptions $USAMt ($isa $X $Person)) which means: Let X be a person, throughout context $USAMt...One could assert in the $USAMt context ($citizenOf $X $USA) and - if there were no further restrictions on $X - that would mean that in that context every person mentioned in any way, is by default, a US citizen.' .[253]

As the illustration suggests, the default settings for a context in the Lets dimension can be considered as antecedents of the various statements that are asserted in that context.

---

[251] Cfr. D. B. Lenat, DCS, p. 62. On the same page Lenat states and describes the two CYC predicates for representing Justification as follows:

'($mtSource C SOURCE) means that the information in context C originated with, or is authenticated by SOURCE - a temporal thing which must either be an instance of $Agent or of $InformationBearingThing.

($mtSourceType C STYPE) means that the information in context C originated with, or is authenticated by, an information source of type STYPE - and $ExistingObjectType which is a subset of the collection $TemporalThing (usually is moreover a subset of $Agent or a subset of $InformationBearingThing.'

[252] Cfr. D. B. Lenat, DCS, p. 63.

[253] D. B. Lenat, DCS, p. 63.

3.4. Context specification and inference in the new strategy

Lenat upholds that, considered as a region in 12 dimension space, setting up contexts in CYC is greatly simplified. In the previous strategy, each context was considered a reified term in the CYC vocabulary. To set up a context, the knowledge enterer was faced with the difficult task of finding, correcting and formulating when necessary, useful assertions and their underlying assumptions. This procedure was facilitated using copy/edit functions to lift assertions and assumptions from one context to another. As more contexts were created it became increasingly difficult to find the most adequate context for new assertions and consistency errors were introduced as a result of lifting. These problems made it necessary to rethink how contexts could be incorporated in CYC. Lenat upholds that these problems can be largely overcome in the new strategy.

In the new strategy, a context is not necessarily a reified term in the CYC vocabulary. Lenat considers it akin to a large non-atomic term (NAT). It is a matrix of settings along the dimensions that are relevant for that context, in which most of the settings can be selected from predefined options. Having 12 dimensions, however, several of which have various sub-dimensions, means that the number of matrix slots for specifying settings is somewhat overwhelming. Lenat explains, however, that task of specifying a context is not as monumental as it might appear in mathematical terms. The benefits that come with the new strategy for specifying contexts is based on two fundamental points which Lenat expresses as follows:

> 'Point 8: One conceptual operation that needs to be supported efficiently is: given an assertion, return the contexts (the set of 12 settings) in which it holds.
>
> Point 9: One conceptual operation that needs to be supported efficiently/naturally is: given a particular context, modify one (or more) of those 12 dimension settings a little.'[254]

Lenat proposes that once these two conceptual operations are efficiently supported, then there are five principle ways that a context for one or more assertions can be specified. In the first case, which Lenat considers the majority, another assertion that is true in the same context or in a similar context can be easily found using advanced browsing tools similar to the ones that are currently used in CYC. Once located, its 12-settings can be examined, and the setting that are relevant can be adopted for the new context. In the second case, the settings from two or more assertions need to be merged in order to establish the required settings for the new context. Lenat explains that in this case settings are not directly modified. The principle activity consists in selectively adopting settings, indicating when the specification to be adopted in the new context is the result of the conjoining, disjoining, intersection or union of the specifications from

[254] D. B. Lenat, DCS, p. 66.

the source assertions.[255]

The third principle case in which a context for a group of assertions is specified is the case where one or more settings need to be adjusted in order to adequately represent the desired new context. Lenat explains that this process can be facilitated by providing intuitive and efficient interface tools for adjusting specifications. In the fourth case, the context to be defined is specialized and unique to the extreme that it is preferable to build it from zero. Lenat claims that these cases are not as difficult and time consuming as they might appear at first. On one hand in may not be necessary to define settings for several dimensions. Furthermore, for several dimensions, it will often be possible to copy settings from contexts in other parts of the KB. Finally, much of the information that is needed to define settings is provided by the terms, antecedents, verbs, structure, etc. in the assertions for which the unique context is specified.[256] The fifth case that Lenat describes regards the situation where settings can copied from reified contexts in CYC. Lenat explains that at times it may be necessary to reify a context in order to facilitate further data entry and inference later on. Instead of seeking out assertion and thereby indirectly its context, one could refer directly to a reified context by name.

In addition to the gains for knowledge entry, Lenat is confident that his new strategy will speed up inferencing. His key point is that, configured along 12 dimensions, CYC can use contexts in a more intelligent manner to quickly find solutions to common sense problems. He writes:

> 'Increasing constraints, dimensionwize, should continually reduce the set of assertions to worry about - sort of like conjoining multiple filters. Alternatively, think of contexts as determining not an absolute relevant/irrelevant black&white border, but rather as determining a partial ordering of relevance, so the most-likely-to-be-relevant assertions can be found and considered before the less likely ones.'[257]

Lenat's point is based on the paradigm of the Knowledge Principle and the Breadth Hypothesis. According to this paradigm, problems solving is fundamentally a search activity. Problems are solved by searching for solutions. Lenat upholds that intelligent problem solving takes place by progressively and intelligently narrowing down the set of possible answers or search space. The search space is reduced by bringing to bear a wide variety of knowledge about the problem and its context. This knowledge acts as constraints or filters on the search space, eliminating irrelevant and unlikely solutions. When sufficient knowledge can be brought to bear, the space can be efficiently and

---

[255] Cfr. D. B. Lenat, DCS, p. 66.

[256] Cfr. D. B. Lenat, DCS, p. 67. On the same page Lenat stresses that the order in which the dimensions are specified is very important. There ought to be some way to select one dimension after another when setting up contexts.

[257] D. B. Lenat, DCS, p. 69.

effectively reduced to a one or more intelligent answers.[258] Lenat suggests that having 12 well defined dimensions along which all contexts are defined will make it easier to formulate multiple and useful constraints on the search space and, in so doing, produce a small set of intelligent responses.

Another benefit for inferencing in the new strategy concerns logical consistency. In the original strategy, logical inconsistencies propagated throughout the KB due to the large amount of lifting that had to be done in order to define new contexts. This slowed inferencing because sophisticated reasoning had to be employed in order to detect and resolve inconsistencies. In the new strategy, the need for sophisticated inference procedures is greatly reduced. The contexts that are created are better adapted to the assertions they contain and it is a lot easier to ensure consistency when the specification of contexts are confined to 12 or less dimension. The result is that, inconsistencies, when they occur, are mostly isolated and the propagation of inconsistencies is greatly reduced.[259]

Inference in CYC is carried out at the heuristic level. This level uses a variety of special-purpose representations and procedures for speedy inference (inheritance for example) and several modules for generating and comparing arguments for and against a given proposition.[260] Performing inference in the 12 dimension world demands heuristics for dimensionalizing the contexts of the problems to be solved. This is a difficult task and it is not quite clear as yet exactly how this will be realized in CYC. Lenat identifies two sorts of questions that should be supported at the heuristic level in order to reason using context in the new approach. The first is that given an assertion P, there ought to be a heuristic procedure that returns the contexts in which P is true, the contexts in which it is not true and the contexts in which it could possibly hold. A related question that ought to be supported is that given a term Z, CYC finds the contexts in which it is present.[261] Together with these and the heuristic that are already present in CYC, additional modules and methods need to be defined which can guide the heuristic process in an intelligent manner. These modules will enable CYC, for example, to determine when to use which context, when a context is insufficient, when a new context ought to be activated or when contexts ought to be enlarged or shifted.[262] These are important issues that are as yet unresolved in CYC. They suggest important shortcomings in Lenat new strategy which will be discussed in greater detail in the following chapter.

---

[258] Cfr. D. B. Lenat and E. A. Feigenbaum, "On the Thresholds of Knowledge", *Artificial Intelligence* 47 (1991), p. 194.

[259] Cfr. D. B. Lenat, DCS, p. 71.

[260] Cfr. R. V. Guha and D. B. Lenat, "CYC: A Midterm Report", *AI Magazine* Fall (1990), p. 36.

[261] Cfr. D. B. Lenat, DCS, p. 69.

[262] Cfr. D. B. Lenat, DCS, p. 70.

Chapter IV

Underlying principles, strengths and weaknesses in CYC

CYC has many strong points, but it also has weaknesses. Throughout the discussion in the previous chapters I have occasionally hinted as to where these may lie. In this chapter the discussion of the strengths and weaknesses of CYC is taken up in greater detail. In the first section of this chapter, the strengths and weaknesses of CYC's methodological tenets are discussed. The first part of the discussion concerns the philosophical underpinnings of the CYC methodology. This permits a more profound evaluation of the strengths and weaknesses of the method of research that Lenat defends.

In the second section of this chapter, the strengths and weaknesses of the theory of intelligence that CYC tests are discussed. The first part of the discussion regards the philosophical suppositions of Lenat's theory of intelligence. This provides the base for a more profound assessment of the theory of intelligence that is adopted in CYC. In the third section I assess Lenat's efforts to represent and cater for contexts in CYC. The evaluation is based on a prior discussion of the philosophical underpinnings of context representation in CYC. In the fourth section the importance of philosophical evaluation in AI is discussed, an activity that Lenat and his colleagues purposely avoid.

4.1. CYC methodology: Principles, strengths and weaknesses

The methodology of AI to which Lenat adheres is summarized in what he and E. Feigenbaum term the Empirical Inquiry Hypothesis or EH. EH was quoted above in section 2.2. To facilitate the discussion, it is quoted here again. The hypothesis states:

> 'Intelligence is still so poorly understood that Nature still holds most of the important surprises in store for us. So the most profitable way to investigate AI is to embody our hypotheses in programs, and gather data by running the programs. The surprises usually suggest revisions that start the cycle over again. Progress depends on these experiments being able to *falsify* our hypotheses. Falsification is the most common and yet most crucial of surprises. In particular, these programs must be capable of behavior not expected by the

experimenter.'[263]

4.1.1. The underlying principles of EH

EH is an adaptation of the empirical method to AI. It matches in particular, the description of the empirical method according to Karl Popper (1902-1994) which is widely accepted in the scientific community.[264] Lenat and Feigenbaum affirm that progress in AI depends on 'experiments being able to falsify' hypotheses and that 'falsification' is the most crucial aspect of their methodology. The term "falsification" that they use is a trademark of Popperian epistemology and progress through falsification is the central paradigm of the empirical method that Popper formulates.[265]

Adapting the empirical method to AI supposes that the central demands of the empirical method can be fulfilled in AI. In the context of the empirical sciences two processes can be identified, an ascending one which leads to the *construction* of theoretical entities such as concepts and models, and another descending, which consists in testing the validity of the theoretical constructs.[266] In the ascending process, scientific activity is characterized by the construction of the *scientific object*. Artigas explains this term as follows:

> 'If we consider real objects with their observable properties, it would not be possible to obtain mathematical formulations such as those in mechanics. To obtain these formulations it is necessary to construct an object that is fictitious, which substitutes the real objects. This is what we will denominate the *scientific object*.'[267]

The theories and models that make up the scientific object can be constructed in many different ways. There are, however, two basic characteristics of *all* models: reference and meaning.[268] Artigas explains this as follows:

> 'Models refer to aspects of reality: for example, the particles of kinetic theory pretend to represent real components of gases, although they do this in a simplified way. Similarly, it is supposed that the corpuscular and frequency

---

[263] D. B. Lenat and E. A. Feigenbaum, "On the Thresholds of Knowledge", *Artificial Intelligence* 47 (1991), p. 204.

[264] For all that refers to the method of scientific investigation in general I rely upon the works of Mariano Artigas. I have worked closely with Artigas over the last two years and I am deeply persuaded by his vision. I have thus chosen to adhere to his work for all that concerns the nature of scientific activity.

[265] Cfr. M. Artigas, *El desafío de la racionalidad*, Pamplona: Eunsa, 1994, pp. 47-52

[266] Cfr. M. Artigas, *La filosofía de la ciencia experimental,* p. 111.

[267] M. Artigas, *La filosofía de la ciencia experimental,* p. 112.

[268] Cfr. M. Artigas, *La filosofía de la ciencia experimental,* p. 119.

> descriptions of subatomic particles correspond to real properties. A similar thing occurs with any model. Nevertheless, the reference to reality is conditioned by the manner in which the characteristics of the model are defined. What is properly studied by the theory is an idealized model, whose characteristics are defined by means of theoretical concepts. As a result, the model has a meaning that is theoretically determined, and a reference to reality whose value ought to be judged by comparing the consequences obtained from the model with the results of experimentation.'[269]

Artigas insists that it must always be possible to evaluate the references in models that are proposed in the empirical sciences. Otherwise, there is no way to relate reality to the basic constructs of the model that are to be evaluated. The result of this is that a model without well defined references cannot be effectively subjected to experimental control.[270] Adapting the empirical method to AI supposes that it is possible to evaluate the references to real intelligence in the hypotheses about intelligence that are proposed. Do Lenat and Feigenbaum adequately account for this? Do the basic constructs of their models have well defined references which subsequently permit experimental control?

The formulation of the Empirical Inquiry Hypothesis (EH) indicates that, in the view of the authors, AI is about investigating natural intelligence. In their description of EH, the two AI veterans refer to "Nature" as holding the most important surprises for us. AI has the objective of uncovering the secrets of intelligence as it occurs in nature. Introducing the Empirical Inquiry Hypothesis, Lenat and Feigenbaum affirm: 'We scientists have a view of ourselves as terribly creative, but compared to Nature we suffer from a poverty of the imagination; it is thus much easier for us to uncover than to invent.'[271] The authors refer to themselves as scientists and reveal that their work is about explaining natural phenomena. This claim suggest that the models they propose refer to real intelligence. How is the reference established? How well defined is the reference?

The Empirical Inquiry Hypotheses has the peculiar characteristic that the reference is established through the mediation of digital machines. The hypothesis explicitly states that: 'the most profitable way to investigate AI is to embody our hypotheses in programs, and gather data by running the programs.' The elements of the theories and models of intelligence that make up the hypotheses are referred, in the first instance, to properties of computer systems. AI, however, as Lenat and Feigenbaum claim, pretends to investigate real or natural intelligence. The empirical method, as described above, demands of AI a reference to natural intelligence. How can hypotheses and models that refer, in the first instance to realities in computer systems, refer also to natural intelligence?

In their formulation of EH, the authors demand that the programs that are tested be

---

[269] M. Artigas, *La filosofía de la ciencia experimental,* p. 119.

[270] Cfr. M. Artigas, *La filosofía de la ciencia experimental,* p. 120.

[271] D. B. Lenat and E. A. Feigenbaum, OTK, p. 204.

capable of behavior that is not expected by the experimenter. There must be "surprises". Lenat and Feigenbaum describe what they mean when they speak of surprises as an essential requirement. They explain:

> 'What do we mean by "a surprise"? Surely we wouldn't want to increase surprises by having more naive researchers, less careful thought and planning of experiments, sloppier coding, unreliable machines, etc. We have in mind astronomers getting surprised by what they see (and "see") through telescopes; i.e. things surprising to the professional.'[272]

The analogy that the authors establish here between the observations of an astronomer and the observations obtained from running a computer program is also surprising. An astronomer, by means of a normal telescope, observes nature directly. An AI researcher, however, observes data obtained by running a computer program. Lenat and Feigenbaum suggest that the programs must be written in such a way that they refer to how natural intelligence works. Just as the astronomer is surprised when he observes some unexpected natural phenomena, the programs in AI ought to reflect natural intelligence in such a way that the AI researcher encounters unexpected results in a surprising way. Lenat and Feigenbaum suggest, implicitly, that it is correct to suppose a parallel between how an appropriately programmed computer works and the dynamics of natural intelligence. When they describe their methodology, however, the authors neglect to illustrate how this parallel is articulated in the process of experimentation.

In AI, a parallel between how a computer works and natural intelligence is embodied in the physical symbol systems hypotheses.[273] In a section prior to their discussion of EH, Lenat and Feigenbaum recognize that this hypothesis underlies research in AI. They cite the symbol systems hypotheses as the underlying principle of their research paradigm. The authors write:

> 'Half a century ago, before the modern era of computation began, Turing's theorems and abstract machines gave a hint of the fundamental idea that the computer could be used to model the symbol-manipulating processes that make up that most human of all behaviors: thinking.
>
> Thirty years ago, following the 1956 Dartmouth Summer Conference on AI, the work began in earnest. The founding principle of the AI research paradigm is really an article of faith, first concretized by Newell and Simon:
>
> **Physical Symbol System Hypothesis**. The digital computer has sufficient means for intelligent action; to wit: representing real-world objects, actions, and relationships internally as interconnected structures of symbols, and applying symbol manipulation procedures to those structures.'[274]

The physical symbol system hypothesis establishes that thinking, like computer

---

[272] D. B. Lenat and E. A. Feigenbaum, OTK, p. 204.

[273] Cfr. Section 1.2.3.

[274] D. B. Lenat and E. A. Feigenbaum, OTK, p. 193.

operations, consists fundamentally in symbol manipulation. The hypothesis concludes that a machine is capable of thought. Through the prism of the Symbol Systems Hypothesis, Lenat and Feigenbaum go on to observe that the AI projects that have had most success when put to the test, are those that embody a knowledge intensive model of intelligence. They avoid explaining this important hypothesis in more detail perhaps to evade philosophical complications.[275] Returning to the reference problem, the gap in Lenat and Feigenbaum's description may find an explanation here. The authors propose a method where the immediate references of the models are to realities in computer systems. They go on to claim that they investigate real intelligence without clarifying how models that refer to computer systems can serve to investigate real intelligence. Here lies a large and problematic gap in the methodology they propose. It may well be that they tacitly suppose that an adequate reference is provided for by means of the Physical Symbol Systems Hypothesis. In the quote above they refer to the Physical Symbol Systems Hypothesis as 'an article of faith'. The reference established through it is thus weak and uncertain.

Lenat and Feigenbaum, in their outline of the Physical Symbol Systems Hypotheses, state that a digital computer has 'sufficient means' for intelligent action. Actual intelligent action is displayed when the computer has been appropriately programmed. Empirical inquiry demands a way of evaluating the extent to which the computer has been appropriately programmed. How do Lenat and Feigenbaum meet this demand? What criteria do they use to judge the extent to which the program embodies an adequate model of genuine, natural intelligence? Considering how Lenat was led to CYC, an important factor for them is the ability to perform well in several domains. Lenat abandoned his AM and Eurisko programs because, though they ran in several domains, their ultimate limitation was their incompetence in a wider selection of domains.[276] Another factor that the authors stress is the ability to surprise. Lenat and

---

[275] Lenat and his colleagues at Microelectronics and Computer Consortium (MCC) where CYC began generally avoid philosophical questions. They see their principle activity as constructing and testing models of intelligence. In the introduction to this work I referred to a statement by Lenat in which he refers to the ontology that philosophers have done for millennia as mere 'theorizing' (See page 1). In their book on Artificial Intelligence, Elaine Rich and Kevin Knight, colleagues of Lenat at MCC, openly manifest their unwillingness to address philosophical issues. Speaking on the relationship of AI with philosophy they declare: 'Philosophy has always been the study of those branches of knowledge that were so poorly understood that they had not yet become separate disciplines in their own right. As fields such as mathematics or physics became more advanced, they broke off from philosophy. Perhaps if AI succeeds it can reduce itself to the empty set.' (E. Rich and K. Knight, *Artificial Intelligence*, New York: McGraw-Hill, 1991, p. 3).

[276] Cfr. D. B. Lenat and E. A. Feigenbaum, OTK, p. 206.

Feigenbaum affirm with conviction that programs which are incapable of producing surprises ought to be discarded as candidates for authentic AI research. The authors write:

> 'The inverse of EH is cruel: If one builds programs which cannot possibly surprise him/her, then one is using the computer either:
>
> (a) as an engineering workhorse, or
>
> (b) as a fancy sort of word processor (to help articulate one's hypothesis), or
>
> (c) as a (self-) deceptive device masquerading as an experiment.'[277]

As an example of what is not AI, they describe their work on the PUP5 program to generate a learning program that Patrick Wilson had written for his thesis some years earlier. They had a well defined target, Wilson's program, and they worked towards meeting this objective. They devised an idealized dialogue which was interpreted perfectly by the synthesizer, generating in clean LISP code the target version of Wilson's program. The authors conclude:

> 'There was not much else PUP5 could do, therefore, besides hit its target, and there was not much that we learned about automatic programming or intelligence from that six month exercise. There was one crucial meta-level lesson we did learn: You can't do science if you just use a computer as a word processor, to illustrate your ideas rather than test them.'[278]

PUP5 had a narrow well defined target and was incapable of producing surprises. Lenat learned that authentic AI investigation ought to investigate models that have no particular target behavior in mind. He embarked on the AM and EURISKO projects in which heuristics guided searches to find solutions. There were no preconceptions coded in about what answers ought to be given. These programs, Lenat claims, produced many surprises and did interesting things. Evaluating these results, Lenat discovered that the ultimate limitation was not CPU or the need to learn new representations, but rather the need to have a large fraction of consensus reality.

Lenat's itinerary in AI research led him to the conviction that a criteria for true empirical investigation is that there be surprises. Together with this, as discussed above, success depends on the ability to perform well in a wide selection of domains. If CYC gives rise to surprises that are more sophisticated in terms of the intelligence required, and its performance in a wide selection of domains is better than that of previous systems, then the CYC representational strategy is a step ahead for building intelligent systems and understanding cognitive processes.

Considered in this way there is an important supposition in the approach that Lenat and Feigenbaum endorse. The authors suppose that a computer system that

---

[277] D. B. Lenat and E. A. Feigenbaum, OTK, p. 205.

[278] D. B. Lenat and E. A. Feigenbaum, OTK, p. 206.

demonstrates what is considered to be intelligent conduct (interesting surprises and performance in a wide selection of domains) can be considered, to some degree, really intelligent. This explains the insistence of the authors that there must be surprises and that the system ought to perform well in many domains. If intelligence were mostly independent of surprises and performance in many domains, then such conduct would be an unimportant, trivial and poor sign that there is some degree of intelligence. Seeking such conduct would surely not serve as a strategy of investigation for a project of the scale of CYC which is projected to consume 2 person-centuries of time and 50 million US dollars.[279]

The underlying supposition indicated above explains the apparent boldness in Lenat and Feigenbaum's discussion of EH which Brian Smith brings to light. Smith observes that the authors speak of using computers as a "tool", the way astronomers use telescopes.[280] He is surprised by this claim, because a telescope is used to directly investigate natural phenomena whereas in AI research computers test models whose adequacy and usefulness must be subsequently evaluated. Smith writes:

> 'For astronomers, telescopes are *tools*, not *subject matters*; the theoretical notions in terms of which we understand telescopes are not the constitutive notions in terms of which we understand *what is seen through telescopes*. AI, in contrast, is different: we exactly *do* claim that computational notions, such as formal symbol manipulation, *are* applicable to the emergent intelligence we computationally model.'[281]

He is perplexed by Lenat and Feigenbaum's analogy because it seems to stray from what in fact, Lenat and Feigenbaum have been doing for years and continue to do on a daily basis. The authors, in practice, devise and test models, an activity which is quite different from using a computer as a tool, the way astronomers use telescopes. Smith concludes: 'The issues here are so complex it is hard to tell what they think: at best they seem to have in mind what would normally be called hypothesis testing, not empirical inquiry.'[282]

---

[279] Cfr. D. B. Lenat and E. A. Feigenbaum, OTK, p. 210.

[280] Describing some of the consequences of EH, Lenat and Feigenbaum write:

'AI is a science when we use computers the way Tycho Brahe used the telescope, or Michaelson the interferometer - as a tool for looking at nature, trying to test some hypothesis, and quite possibly getting rudely surprised by finding out that the hypothesis is false. There is quite a distinction between using a tool to gather data about the world, and using tools to, shall we say, merely fabricate ever more beautiful crystalline scale models of a geocentric universe.' (D. B. Lenat and E. A. Feigenbaum, OTK, p. 188).

[281] B. C. Smith. "The owl and the electric encyclopedia" *Artificial Intelligence* 47, 1991, p. 257.

[282] B. C. Smith. "The owl and the electric encyclopedia", p. 258.

The underlying assumption at the heart of Lenat and Feigenbaum's proposal explains the apparent dichotomy between theory and practice in their proposal, which Smith observes. The two authors suppose that a computer system that demonstrates what is considered to be intelligent conduct (interesting surprises and performance in a wide selection of domains) can be considered, to some degree, to be really intelligent. In harmony with this principle, the authors suppose that in an advanced stage of AI, the surprises and performance will be sophisticated and powerful to the extent that computers become subjects of real human level intelligence and beyond. When Lenat and Feigenbaum speak of using a computer as a tool they refer in greater part to this advanced stage of AI and not to their day to day activity which concerns the building of intelligent systems. In an advanced stage computers become tools because, like telescopes, their role is to gather data which serves to understand an objective reality, in this case, the computer's intelligence. Lenat and Feigenbaum concede that AI is still leagues away from its mission. For this reason AI as a science is far from maturity.

4.1.2. CYC's methodological strengths

The benefits that come from EH are similar to those that derive from the empirical method of research in general. The process of falsification that the empirical or scientific method entails guarantees a measure of progress. New hypotheses are formulated to eliminate the errors encountered when previous hypothesis are empirically evaluated. The experiments are well defined and repeatable. They provide a context which permits the scientific community to objectively determine whether a new hypotheses represents an advance or not over older theories.[283] Adhering to EH, CYC presents itself as building scientifically on the experience of AI. This is an important strength because it suggests an objective advance in AI were it to give better results (new and more interesting surprises together with improved competence in a wide selection of domains). CYC separates itself from isolated efforts to build intelligent machines based on ingenious theorizing. Replying to an accusation that their project is hasty and too ambitious, Lenat and Feigenbaum write: 'If we fail, then the next set of important lessons for AI are likely to emerge by tackling this large empirical task, rather than by micro-experiments or sterile philosophical argument.'[284] The project draws on the experience of the past and EH protects it from theoretical critiques that have not been shown to be relevant empirically. From this point of view, that is, as a way to establish inter-subjectivity and to guarantee progress, EH is an important strong point of CYC.

4.1.3. CYC's methodological weaknesses

While EH may strengthen the relevance of CYC in terms of objectivity and

---

[283] Cfr. M. Artigas, *La filosofía de la ciencia experimental,* pp. 135-142.

[284] D. B. Lenat and E. A. Feigenbaum, "Reply to Brian Smith", *Artificial Intelligence* 47 (1991), p. 235.

progress, several weaknesses arise when EH is carefully confronted with the demands of empirical investigation which EH proposes for AI. Adapting the empirical method to AI supposes that it is possible to evaluate the references to real intelligence in the hypotheses about intelligence that are proposed. Lenat and Feigenbaum fail to explain clearly how this is articulated. They suppose that once there is sufficient meta-level knowledge to fall back on, and appropriate heuristics are provided, then a computer system is capable of real intelligent action. This underlying stance suggests that the authors rely on the Physical Symbol Systems Hypothesis for the reference to real intelligence that the empirical method requires. The hypothesis claims that it is correct to affirm a parallel between how an appropriately programmed computer works and the dynamics of natural intelligence because both are symbol systems.

How symbols and representations are used by humans is still poorly understood. In fact, there are several approaches in AI today that challenge the Physical Symbol Systems Hypothesis. In the first chapter, two such approaches were described, replacement connectionism and anti-representationalist robotics. Researchers in replacement connectionism consider the hypothesis untenable because it ignores the physical properties and organic operations of the brain. They uphold that intelligence emerges from the patterns of connectivity between neurons when intellectual activity takes place. They insist that the patterns that are produced are neither representations nor symbols. Intelligence is an emerging and evolving phenomena which cannot be represented using formal structures.[285] In anti-representationalist robotics, the underlying conviction is that intelligence emerges from lower-level processes that arise when autonomous agents interact with the real world. Researchers in this area consider the Physical Symbol Systems hypothesis presumptuous and misleading because it overlooks many important aspects of human cognition such as situatedness, embodiment and emergence.[286]

The difficulties raised by researchers in connectionism and anti-representational robotics seriously challenge the plausibility of the Physical Symbol Systems Hypothesis. Added to this, the Chinese room scenario proposed by John Searle illustrates that there is no reason to suppose that symbol manipulation explains understanding. Searle demonstrates that as long as a program is defined in terms of computational operations that manipulate formally defined elements, then neither those elements nor their manipulation have interesting connections with understanding. Searle's point of view is that programmed computers following formal instructions understands what a car or an adding machine understand: exactly nothing. The computer's understanding is zero.[287]

New approaches in AI and Searle's argument suggests that the Physical Symbol

---

[285] Cfr. section 1.2.2.

[286] Cfr. section 1.2.3.

[287] Cfr. section 1.3.1.

Systems Hypothesis is misleading. Lenat and Feigenbaum rely on a poor and conjectural support for their reference to real intelligence. An important weakness in the methodology that they propose is that according to the same empirical method they uphold, their results will always be of little use for explaining real intelligence.

Linked to their reliance on the Physical Symbol Systems Hypothesis, empirical inquiry demanded of the authors a way of evaluating the extent to which the computer has been appropriately programmed. It was observed that the authors suppose that a computer system that demonstrates what is considered to be intelligent conduct (interesting surprises and performance in a wide selection of domains) can be considered, to some degree, really intelligent. In harmony with this principle, the authors suppose that in an advanced stage of AI, the surprises and performance will be sophisticated and powerful to the extent that computers become subjects of real human level intelligence and beyond. This supposition is clearly an adaptation of the Turing Test to the project they propose. Turing was very demanding. In the test he proposed the computer system ought to be sufficiently advanced so as to escape identification as a computer system most of the time. Lenat and Feigenbaum suppose an immensely simplified version more in accord with the models and techniques at their disposition. A computer system begins to show intelligence when there are unexpected surprises that suggest intelligence and the system performs well in a wide selection of domains. Their criteria are vague and imprecise, a far cry from what Turing had in mind.[288] In addition to this difficulty the argument raised by Searle continues to apply. The fact that a computer system demonstrates what is considered to be intelligent conduct does not imply that it understands anything. These are important weaknesses in the methodology that the authors propose.

4.2. CYC theory of intelligence: Principles, strengths and weaknesses

In section 2.3. it was explained that the columns of CYC theory of intelligence are the Knowledge Principle, The Explicit Knowledge Principle and the Breadth Hypothesis. Based on these principles, Lenat proposed a three phase project to build and test intelligent systems. The first stage of the project involves hand-coding the essential knowledge items and rules of thumb that comprise consensus reality. Though a great deal of effort has been invested, the first stage is far from completion. The representational scheme and technical methods that have been adopted have been reworked several times in response to difficulties that have been encountered along the way. Furthermore, Lenat's recent paper on context suggests that new and radical revisions are needed. In this section the underlying principles of CYC theory are discussed and the strengths and weaknesses of Lenat's proposal are evaluated.

[288] Cfr. section 1.3.1.

4.2.1. Underlying principles in CYC theory

CYC is best understood as a scientific experiment. The theory of intelligence that CYC tests is fruit of a specific methodology which Lenat and Feigenbaum describe in the Empirical Inquiry Hypothesis. This hypothesis places important demands on how new theories ought to be formulated. EH establishes that AI investigation is driven by a process of falsification. The hypotheses are embodied in programs and the programs are run to obtain data. This data is evaluated according to the sophistication of the surprises and the level of performance in a wide variety of domains. During testing important limitations are observed. These limitations suggest ways that the original hypotheses can be improved. EH stipulates that the new hypotheses that are proposed *must* address the shortcomings that were encountered while safeguarding prior achievements. CYC theory is a product of this dynamic and it is in this context that it ought to be understood.

Lenat's itinerary and thought leading up to CYC illustrates how the rigorous application of EH gave rise to the project. In the previous section is was explained that Lenat was led to CYC due to shortcomings in the AM and EURISKO projects that he had previously pioneered. The results of these experiments suggested the need for a large storehouse of general knowledge. Lenat also observed that knowledge-based systems enjoyed a great deal of success compared to other types of systems. This suggested to him that his diagnosis was well aimed. Lenat detected, however, that an important shortcoming in expert systems was their brittleness. Performance dwindled when they were faced with novel situations. To advance in AI, EH demanded the formulation, elaboration and testing of new hypotheses and models of intelligence that are knowledge based and which overcome brittleness. The Knowledge Principle, Explicit Knowledge Principle and the Breadth Hypothesis are Lenat's response to this challenge.[289] Having established these principles, Lenat proposed a three phase program for building and testing his hypotheses. The first phase of the project is to make explicit and hand code a huge KB containing the rules and knowledge items that comprise common sense. Lenat and his team have devised detailed technical proposals for

---

[289] Cfr. section 2.3. The Knowledge Principle affirms that a system exhibits intelligent understanding and action at a high level of competence primarily because of the *knowledge* that it can bring to bear: the concepts, facts, representations, methods, models, metaphors, and heuristics about its domain of endeavor. The Explicit Knowledge Principle asserts that for knowledge to be useful, it needs to be explicitly defined. The essential characteristics of the knowledge items ought to be specified in a consistent way and it ought to be clearly indicated how the knowledge items relate to each other. Finally, the Breath Hypothesis is a mandate to build systems that incorporate two fundamental strategies for dealing with novelty. On one hand, the system must have the capacity to fall back on increasingly general knowledge and secondly, the system ought to possess the ability to reason by analogizing to specific knowledge from far-flung domains.

representing and storing such information, for actualizing the KB and for performing inference.

As a product of EH, Lenat's proposal is grounded in the suppositions at the root of EH which were discussed in the previous section. The authors suppose that the Physical Symbol Systems Hypothesis is correct and they rely on a weak version of the Turing test for evaluation. These suppositions help to explain many of the deficiencies in Lenat's theoretical framework which Brian Smith points out. Smith characterizes Lenat and Feigenbaum's proposal as enthusiastic and presumptuous. He accuses the authors of moving from broad intuition to detailed proposals, assuming away the intermediate conceptual problems that are of central concern in AI. Commenting this defect, he writes:

> 'One is reminded of tunnel diodes. For a moment the argument is on the plane of common sense, and then it is suddenly at an extreme level of specificity, without ever having been anywhere in between. From the generality of human knowledge to the intricacies of slot inheritance; from the full flowering of intelligence to particular kinds of controlled search. The problem is not simply that the reader may disagree with the conclusions, but that there is no hint of the complex intellectual issues and decades of debate that lie in the middle. Whereas tunneling electrons - or so we are told - genuinely switch from one place to another without ever being half-way in between, arguments do not have this luxury.'[290]

After presenting his accusation in general terms, Smith goes on to examine in greater detail some of the important issues that Lenat and Feigenbaum fail to address. Concerning the Knowledge Principle, Smith identifies a missing middle realm between the statement of the principle and the mandate to construct a large KB of consensus reality drawing on the experience of expert systems. He argues that the authors brush aside fundamental questions about the notion of knowledge such as how it is acquired and used, the role of perception and motor coordination and the phenomenon of tacit expertise.[291] Regarding the Breadth Hypothesis, Smith maintains that Lenat and Feigenbaum fail to illuminate the subtleties of analogy which the hypothesis prescribes as a general strategy. Smith explains that a proper treatment of analogy requires a notion of *relevant* similarities along with other deeper issues that have been investigated for years but which the authors assume away. In addition to this, the Breadth Hypothesis recommends falling back on more general knowledge as an approach for finding solutions. Smith considers this a clamorous over simplification. He complains that falling back on more general knowledge seems to resolve everything: pragmatic assumption, concept formation, inference, induction over experience, formations of judgment, theory change, discourse understanding and other such phenomena which are still poorly understood.[292] Smith argues that the authors present the Breadth Hypothesis

---

[290] B. C. Smith. "The owl and the electric encyclopedia", p. 253.

[291] Cfr. B. C. Smith. "The owl and the electric encyclopedia", p. 255.

[292] Cfr. B. C. Smith. "The owl and the electric encyclopedia", p. 257.

in such a way that it seems to automatically cater for all these phenomena and they do not bother to provide the theoretical details of how this can come about. What is even more striking for Smith, is the suggestion that all this knowledge can be captured in a million frames.[293] Smith considers the theoretical framework that is proposed for a CYC a case of premature formalization, lacking an important middle realm.

The tendency in CYC to formalize prematurely and to assume away important phenomena that are still poorly understood is rooted in the methodology that is adhered to and the underlying assumptions in CYC. EH, as Lenat and Feigenbaum explain, is a mandate to build systems. It invites researchers to proceed step by step, building on previous accomplishments. In such a context of investigation, simplifying assumptions are necessary in order to favor experimentation and to center the investigation around well defined problems that build incrementally on prior achievements. Considered in this way, to Smith's outrage, many of the complicated issues that he raises ought to be dealt with at a later stage and be assumed away for present purposes. Lenat and Feigenbaum propose that, adhering to EH, research in AI ought to center around the building of a large knowledge base of common sense knowledge, putting aside issues and theories, which, at present, are poorly understood.

The empirical method which Lenat and Feigenbaum adopt not only favors a knowledge approach, but establishes theoretical standards for the models that are proposed within such an approach. New models ought to cater for the shortcomings in previous models while conserving advances that have been achieved. The principle point of reference for establishing theoretical standards are the achievements of prior systems and not the degree of intellectual or logical rigor that is displayed. Discussing this aspect of the empirical method, Artigas explains:

> 'The greater the number of consequences of different types that can be experimentally confirmed, the more we can trust in the hypotheses from which they are deduced, above all if the predictions are precise and previously unknown.'[294]

Artigas goes on to describe five criteria that are used in practice. These criteria describe in detail how theoretical standards are elaborated and applied in empirical inquiry.[295] New theories ought to have the capacity to give rise to an overall improved

---

[293] Cfr. B. C. Smith. "The owl and the electric encyclopedia", p. 257.

[294] M. Artigas, *La filosofía de la ciencia experimental,* p. 138.

[295] The first criteria is the explicative power or the capacity of the hypotheses to explain the problems that have been identified and the data that has been gathered. The second criteria is the predictive power which is the capacity of the hypotheses to make correct predictions. The third criteria is the precision of the explications and predictions. The fourth criteria concerns the compatibility of the hypotheses with other varied and independent phenomena. Finally, the fifth criteria is mutual support with

evaluation compared to existing theories. This is the principle theoretical standard and the minimum requirement.

Adhering to EH, Lenat and Feigenbaum understand theoretical standards in this way. Their principle criteria for evaluation is based on a weak version of the Turing test. There ought to be more interesting surprises and improved performance in a wide variety of domains. In the context of EH, the minimum theoretical requirement for new theories is the capacity to produce more interesting surprises and to perform better in many domains compared to expert systems and to the previous projects of the authors. Given that current systems are mostly brittle as Lenat and Feigenbaum observe, the minimum requirements do not demand elaborate theories that account for the full gamut of intelligent behavior as Smith seems to demand. A simple hypotheses which promises new surprises and improved performance in many domains meets the minimum theoretical standards. Understood in this way, the proposal that Lenat and Feigenbaum put forth is a valid and even good response to the challenges of EH.

In his evaluation of the CYC proposal, Smith uses theoretical standards that are based on intellectual and logical rigor rather than experimental results. It is not surprising that he accuses the authors of premature formalization and of assuming away an important middle realm. In their reply to Smith, Lenat and Feigenbaum point out this misunderstanding. The authors write:

> 'Smith implies that we believe that all the theoretical foundations of AI will be complete by 1994. We certainly do not believe that. In fact, a host of fundamental research questions may be uncovered by this work, and become seen as important...The same situation occurs when high energy experimental physicists gather data about collisions at new energy levels, etc. etc. It is foolish never to theorize, but it is commonplace for empirical experiments and constructions to outstrip (and drive) the development of theory, especially in a field's first few centuries of life. In its early stages, a theory may be little more than a plausible generalization of a class of recently observed phenomena. Theory building must - and does - go on in the absence of complete sets of data to characterize; and experiments must - and do - go on in the absence of complete theories.'[296]

The three stage project that Lenat proposes is a strategy to build an intelligent system without a complete theoretical understanding of what cognition entails. Furthermore, as AI is still in it early stages, theories that are little more than plausible generalizations are acceptable for testing. The results of these tests will drive further theory building and testing. Lenat and Feigenbaum go on to reaffirm that in the current state of AI, the maximum gain can be obtained by attempting to build and test intelligent systems without having clear answers to some of the questions that Smith

---

other established theories (Cfr. M. Artigas, *La filosofía de la ciencia experimental,* pp. 138-142).

[296] D. B. Lenat and E. A. Feigenbaum, "Reply to Brian Smith", *Artificial Intelligence* 47 (1991), p. 232.

raises. They suggest, for example, that the problem of "genuine semantics" will get easier as the KB grows and they openly admit that they use symbol structures to represent things without understanding exactly how sets of symbols relate to the world.[297] Adhering to a different sort of theoretical standard, Smith regards these problems as part of the middle realm that ought to be elaborated in the theoretical models that are proposed in AI.

4.2.2. CYC's theoretical strengths

The hypotheses, models and methods that make up CYC theory have many strong points. Understood in the context of EH, they constitute an acceptable and even good response to the demands of EH. The Knowledge Principle is a simple, clear and well defined axiom that builds on the experience of AI and prior experimental results. It is a powerful principle because it applies in all aspects of intelligent activity and indicates that priority should be given to knowledge based approaches in solving problems.

In order to test their hypothesis, the authors consider that the best way to proceed is by building a large knowledge base of consensus reality. They choose to represent knowledge explicitly, taking into account the programming tools at their disposition, their relative costs and ease of use. The essential characteristics of the knowledge items ought to be specified in a consistent way and it ought to be clearly indicated how the knowledge items relate to each other. The authors argue that explicit representation makes it easier to update the knowledge base when new information is provided. Furthermore, given a reality to represent, they uphold that an explicit representation is likely to be more usable and useful for other systems.[298] Finally, using explicit knowledge responds to the demand of EH that new systems ought to display surprising behavior. When knowledge is compiled away in the program code the behavior of the system is entirely predictable. On the contrary, when knowledge is explicitly represented, the knowledge items are independent of the program code. They can be updated and modified in dynamic and interesting ways, giving rise to surprising results. In the context of EH, the explicit approach chosen by the two AI experts has many advantages and is an acceptable way of proceeding.

To address the problem of brittleness, Lenat and Feigenbaum propose the Breadth Hypothesis. This hypothesis is in harmony with the Knowledge Principle and the explicit knowledge approach that they adopt. It adds to these principles a strategy for dealing with novelty - broad knowledge and analogizing to far flung domains. Smith considers these strategies to be oversimplifications. In the context of EH, however, the Breadth Hypothesis is an acceptable proposal. It provides a simplified and direct

---

[297] Cfr. D. B. Lenat and E. A. Feigenbaum, "Reply to Brian Smith", p. 240.

[298] Cfr. D. B. Lenat and E. A. Feigenbaum, "Reply to Brian Smith", p. 242.

response to a central shortcoming in previous models, that of brittleness. The hypotheses articulates with the other principles in a commendable manner and it lends to embodiment in programs and testing.

Together with the strengths that have been indicated above, an important strength of CYC is the formidable upper ontological scheme of over 3,000 terms that has been elaborated and implemented over the years. Lenat claims that the scheme is universal and articulate and that it has been tested using millions of examples.[299] As a result of this, other systems can share the same upper ontology with CYC, even though at a lower level the terms and their organization are entirely different. Systems that share the CYC upper ontology will also be able to benefit from the vast quantity of lower level terms, notions and rules of thumb, that have been entered into CYC and they can exchange information with other applications that share the upper CYC ontology.

Aside from these benefits, the CYC upper ontology contains original and provocative descriptions for notions such as substance, individual, events, actions, agents, spatial and temporal things, intangible things, change, movement, cause, granules and much more. The upper ontology also describes how these terms are related.[300] Lenat stresses that the ontological scheme that he and his team have devised is fruit of ontological engineering and not ontological theorizing. The upper ontology, like the other hypotheses is best understood in the context of EH. It was built by studying what, in fact, are the most general common sense terms that are used in western society, how they are related, and how they can be described so as to be coherently stored in a large KB. Lenat's work lacks philosophical grounding but is nevertheless a useful and interesting reflection of how we commonly think about the world.

4.2.3. CYC's theoretical weaknesses

In the previous section, the strengths of CYC theory were evaluated in the context of EH. In this context the theory that Lenat and Feigenbaum propose is acceptable and in many ways praiseworthy. While this may be so, several of the issues that Smith raises suggests that there are also important weaknesses. Is EH an adequate context for building theories in AI? Given the costs involved and the availability of profound and useful studies about intelligence in both empirical and non-empirical fields, should not CYC theory be subjected to more rigorous theoretical standards that transcend EH?

In the first section of this chapter, it was observed that EH rests on two fundamental assumptions that are mostly conjectural - the Physical Symbol Systems Hypothesis and a weak version of the Turing test that is far from what Turing had in mind. An important weakness in CYC theory is that the context in which it can receive a favorable evaluation rests on two underlying assumptions that are doubtful and vague.

---

[299] Cfr. section 2.4.

[300] Cfr. section 2.4.

In many ways the prematureness and lack of rigor that Smith complains of has its roots here. In the empirical sciences in general, the reference to reality is mediated by instruments that are exact and effective, and the measurements depend on theories that are proven, precise and trustworthy. Furthermore, the criteria for evaluation rest on precise comparisons.[301] This drives theoretical activity, demanding rigor and thoroughness in the new hypotheses that are proposed. In EH, the reference is mediated by a general and uncertain principle and the criteria for evaluation is a vague and weak form of the Turing test. This gives rise to low standards of rigor and thoroughness for the new hypotheses that are proposed. Difficult issues can be assumed away without violating theoretical standards in the context of EH.

The internal difficulties at the heart of EH suggests that it is not an adequate context for building theories in AI. Given the costs involved and the availability of profound and useful studies about intelligence in both empirical and non-empirical fields, should not CYC theory be subjected to more rigorous theoretical standards that transcend EH? Smith answers affirmatively. He writes:

> 'Perhaps someone will object. Lenat and Feigenbaum march to the pragmatist's drum, after all. So it is unfair to hold them to clear theoretical standards? I thing not. For one thing, in a volume on the foundations of AI, explicating premises should be the order of the day. Second, there is the matter of scale. This is a large project they propose - all of consensus reality, 50 million dollars for the first stage, etc. Untutored pragmatism loses force in the face of a task of this magnitude.'[302]

Smith goes on to analyze how CYC theory deals with several foundational questions in AI. He adopts this method as a strategy to uncover and discuss some of the important questions that make up the middle realm which Lenat and Feigenbaum hide away. He draws on the thought and work of important AI researchers, cognitive scientists and philosophers to illustrate the shortcomings in the theory that Lenat and Feigenbaum propose and to defend his own arguments as to where solutions may lie. In the context of more rigorous theoretical standards that transcend empirical inquiry, EH proves to be a poor framework for elaborating hypotheses and models in AI. As a consequence of this CYC theory fails to take into account important phenomena that have been studied and investigated for many years. These are important weaknesses in the theory that is proposed for CYC.

4.3. Context in CYC: principles, strengths and weaknesses

Lenat considers the most important lesson learnt over the years to be the need to incorporate contexts. For over a decade he and his team have invested a great deal of effort to represent and store contexts. Their work helps to further understand and

---

[301] Cfr. M. Artigas, *La filosofía de la ciencia experimental,* pp. 138-142.

[302] B. C. Smith. "The owl and the electric encyclopedia", p. 258.

evaluate their underlying assumptions, motivations and tendencies. In this section the underlying principles of context representation in CYC are discussed and the strengths and weaknesses of Lenat's approach are evaluated.

4.3.1. Underlying principles of context representation in CYC

Context representation in CYC, like CYC theory, is best understood in the context of EH. It is fruit of a process of falsification in which new hypotheses are constructed in response to shortcomings that have been encountered during implementation. Lenat and his team originally proposed to build a flat consistent context-free KB of consensus reality. Under this paradigm, knowledge enterers were required to explicitly define the valid and invalid contexts of use, each time they updated CYC with a new expression. As more items were added it became more urgent to avoid the useless repetition of assumptions for assertions that apply in similar situations. Furthermore, lacking a consistent way to link groups of assertions and assumptions it became more difficult to ensure consistency. An important cause of this difficulty was that knowledge enterers often differed in their appreciation of the contexts and assumptions that ought to be included. Though procedures were put into place to correct this difficulty,[303] Lenat and his colleagues realized that it would be futile to continue in this way. Their experience invalidated their original vision of a flat context-free KB of consensus reality. A new strategy was needed which would avoid the costly repetition of assumptions and guarantee consistency among expressions and assumptions that apply in similar situations.

Around 1990, Lenat and his team began to investigate new ways to get around the difficulties that they had encountered. Guha dedicated his doctoral work to formulating a response to this challenge. Based on his work, a new strategy was adopted in which contexts were represented as reified objects in the KB. Guha called these objects "microtheories". Each context was represented as a series of assertions with a group of associated assumptions. Using the models and techniques that Guha developed for his doctoral project the CYC KB was carved up into hundreds of contexts. By means of sophisticated copy/edit functions, knowledge enterers could create new contexts by copying and modifying assertions and assumptions that already exist in the KB. This process (called "lifting") facilitated data entry and helped to maintain a measure of consistency within each context.

While microtheory technology promised to overcome the early difficulties, important shortcomings arose which undermined and offset its usefulness. As previously explained, lifting turned out to be a bothersome source of logical inconsistencies. Furthermore, as more contexts were created it became difficult and time consuming for the knowledge enterer to select the most adequate context each time

---

303 Cfr. D. B. Lenat and R. V. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project*, Reading, MA: Addison-Wesley, 1990, pp. 30-35.

he/she updated the KB with new information. Moreover, careless knowledge entering was frequently the cause of run time errors which had to be debugged. Finally, Lenat and his team were faced with the difficult challenge of maintaining a good intermediate granularity for the number and choice of contexts. Though microtheory technology provided ways to overcome early difficulties, the shortcoming encountered during its implementation suggested the need for new theories and techniques, starting again the EH cycle.

Lenat's new strategy addresses the shortcomings of microtheory technology. He proposes that context can be adequately defined by specifying values along 12 independent dimensions. Proceeding in this way, Lenat upholds that many of the errors caused by logical inconsistencies among the assertions and assumption in a context can be avoided. In the new strategy, two principle criteria for choosing the dimensions were their utility in separating out mutually-irrelevant portions of knowledge from each other and the ease of computing the overlap or disjointedness of sub-sections of the dimension. This has the effect of separating out inconsistent assertions and assumptions into distinct context regions thereby facilitating consistency. Logical inconsistencies are further limited by having the assumptions for a given context mostly factored out into 12 dimensions. Lenat claims that this ought to reduce the need for special assumptions, making them unnecessary in most cases. It becomes easier to ensure that they are consistent with each other, and that the assumptions are nontrivially assumed by almost all the assertions.[304]

In the previous model, the principle source of logical inconsistencies was lifting. While it proved useful for creating new contexts and updating them with new assertions, the copy/edit functions on which it was based did not adequately cater for consistency. In his new proposal, Lenat explains that contexts will be created and updated in a radically new way. The new strategy for specifying contexts is based on two fundamental conceptual operations that will need to be efficiently supported. In the first place there ought to be interfaces in which given an assertion the system returns the context (the set of 12 settings) in which it holds. Secondly, given a particular context, there ought to be sophisticated tools to modify one or more of the dimension settings.[305] Using these operations, new contexts will mostly be created by adjusting the dimension values of one or more contexts that already exist. Proceeding in this way excludes the introduction of inconsistencies in the 12 major aspects of all contexts. Inconsistencies can arise when specialized assumptions are copied. As such assumptions will be few or non-existent, inconsistencies among them can be easily monitored and controlled.

Driven by EH, Lenat's new proposal also addresses the problem of selecting adequate contexts that stalled data entry in the previous scheme. The proliferation of contexts made it difficult and time consuming for knowledge enterers to choose the

---

[304] Cfr. D. B. Lenat, DCS, p. 71.

[305] Cfr. D. B. Lenat, DCS, p. 66.

most adequate context for new assertions. Lenat claims that knowledge entry becomes easier in the new model. He summarizes his principle arguments for this claim in three points as follows:

> 'Contexts should help speed things up because:
>
> a. In a context, thanks to factored-out assumptions, assertions can be much simpler (terser, shorter, more certain, have fewer exceptions, etc.)
>
> b. Even if the assumptions of a context are only much later - or even never - fully known and specified, it may be relatively easy and natural to tell that assertion P belongs in the same context as assertion Q.
>
> c. The *n* dimensions we choose for context-space provide a set of ways to specify values in those dimensions. Thus, given interface tools that exploit those preconceived ways of specifying regions of context dimensions, it should be possible for a knowledge enterer stating some assertion P to easily broaden/narrow/change the domain over which P is asserted to be true, along those *n* dimensions. I.e., to easily "write" some conjuncts on P's antecedent just by sliding/dialing...'[306]

Points b. and c. depend on conceptual tools and operations that are yet to be designed and studied in greater detail. Lenat refers to sliders, dials, maps, trees, graphs, tables and so on to indicate the sort of technology that might be useful for building such tools. Exactly how such technology can be adapted to his specific needs is an important question that he and his team must address.

The final challenge of EH concerned the granularity for the number and choice of contexts. A good intermediate granularity was required in the previous model in order to boost the effectiveness and efficiency of searching which is at the heart of inference in CYC. Lenat claims that his new strategy overcomes this difficulty. Defining contexts using a small number of dimensions enables the search to be generally limited to small region of relevant contexts when powerful browsing tools are available.[307] Furthermore, inconsistent information is kept isolated in other contexts, thereby reducing the frequency and severity of sophisticated reasoning processes to weigh the pro and con arguments for a proposition.

Lenat's new strategy illustrates that EH continues to guide his research activity and it is in this context that his new proposal ought to be understood. The theory and models that he proposes respond to shortcomings that arose in the previous strategy during implementation. As fruit of EH, however, the deeper underlying suppositions of his strategy remain unaltered. Lenat supposes the validity of the Physical Symbol Systems Hypothesis and relies on a weak version of the Turing test for evaluation. As discussed in the previous section, this has important repercussions on the theoretical standards that

---

[306] D. B. Lenat, DCS, p. 71. Cfr. section 3.4.1. for a more detailed discussions of knowledge entry in the new strategy.

[307] Cfr. section 3.4.

govern the models and hypotheses that he proposes. The theoretical standards that govern Lenat's new proposal is based on the capacity to give rise to an overall improved evaluation compared to the previous strategy. More specifically, as a minimum requirement, new proposals ought to respond to the shortcomings of lifting, the difficulties in choosing contexts for assertions and the problem of maintaining an adequate context granularity in the KB to facilitate searching. A proposal that addresses these issues promises more interesting surprises and improved performance in a wide variety of domains compared to the previous system based on microtheories. It passes the weak version of the Turing test that EH supposes. For his new strategy Lenat continues to adhere to theoretical standards based on EH rather than standards of intellectual rigor and thoroughness which Smith advocated.

4.3.2. Strengths of context representation in CYC

The theory, models and techniques that Lenat proposes in his new strategy have many strong points. Understood in the context of EH, they respond to the demands of EH, catering for shortcoming in the previous model. Lenat takes advantage of the weak theoretical standards which govern the project to simplify his models and theories, assuming away difficult issues that are not demanded by the methodology he adopts. In the first place, he clarifies that his proposal addresses just two of the seven elements of shared context.[308] Lenat explains that the first stage of his project is aimed at storing the facts and rules that we rely on in common sense reasoning. His interest in context at present concerns these two areas. He explains that the other elements of shared context will be dealt with in the later stages of his project. Another important simplification that Lenat makes use of concerns the number of dimensions and how they are chosen. To select the dimension Lenat excluded criteria that go beyond the demands of EH. The criteria that were chosen demand that the dimensions be chosen in such a way that they make it easier to overcome the shortcoming in the previous model. Lenat explicitly excludes considerations that refer to philosophical, linguistic and other models that are not driven by EH. To further simplify his model, Lenat selected out only the first 12 of the numerous categories along which context space can be partitioned. He explains that simplifying the model in this way will provide the maximal utility for the minimal cost at the present stage of research.[309]

According to EH, it ought to be easier to create new contexts and to select the adequate context for new assertions in the new model. Driven by this principle, Lenat simplifies how dimension values are specified and how new assertions are entered. In the time and space dimensions, for example, a difficult issue concerns the representation of temporal and spatial projections.[310] Lenat recognizes that determining projections in a

---

[308] Cfr. section 3.2.1.

[309] Cfr. section 3.2.2.

[310] Cfr. section 3.3.1.

precise manner will stall data-entry. To remedy this handicap, Lenat assumes that one of a handful of predefined types of distributions will be sufficient for the greater majority of cases. CYC can propose options and knowledge enterers will be required to give just a crude estimate of the parameters of the distribution that they select. Default distributions are attached to each context and the assertions that are included automatically inherit these defaults unless an overriding distribution is indicated.

Lenat employs automatic inference, parameter lists and inheritance as a general strategy to simplify the definition of dimension values. Referring to other difficult representation issues in the time dimension Lenat writes:

> 'To make the task of specifying these three types of meta-level information much less time-consuming, each of them will almost always be inferable (i.e. computable) from the predicates and other terms used in the assertion, and/or be a default attached to the context as a whole. In rare cases some special meta-level assertion about A1 will have to be made. A similar tripartite scheme (context-wide unless predicates-involved unless idiosyncratic-for-that-assertion) applies to most aspects of most context-space dimensions, not just Time.'[311]

Lenat suggests that most of the time, meta-level information can be captured without time-consuming and costly keying-in by using inheritance procedures, automatic inferencing, and parameter options. Proceeding in this way most of the information that is worth representing in the KB can be captured efficiently and effectively, avoiding many of the consistency problems that plagued the previous strategy. As a final simplifying assumption, Lenat supposes that sophisticated tools employing dials, sliders, graphs, tables, trees and so on will be available for browsing, search and update. Using such technology demands procedures and structures which may be extremely costly to incorporate into CYC. Furthermore, they may prove inadequate for the complex operations that Lenat demands. Lenat assumes away the difficult questions concerning the design and implementation of these tools on which the success of his new strategy depends. He supposes that the dynamics of EH will resolve or dissolve these issues at some later stage.

Considered in the context of EH, an important strength of Lenat's new proposal are the simplifying assumptions that facilitate data-entry and inference. These assumptions do not violate the demands of EH and they make implementation more feasible.

4.3.3. Weaknesses of context representation in CYC

Lenat's new proposal still fails to address many of the issues which Brian Smith pointed out in 1991. Smith's central accusation was that Lenat moves from broad intuition to detailed proposals, assuming away the intermediate conceptual problems that are of central concern in AI. In his new strategy an important middle realm continues to be missing. Lenat continues to overlook issues that Smith raised such as how knowledge is acquired and used, the role of perception and motor coordination and

---

[311] D. B. Lenat, DCS, p. 29.

the phenomenon of tacit expertise. Other questions still overlooked concern the proper treatment of analogy and aspects of phenomena such as pragmatic assumption, concept formation, inference, induction over experience, formations of judgment, theory change, discourse understanding and so on which are still poorly understood.[312] Lenat supposes that efforts to implement and test his new approach will resolve some of these issues, or at least suggest the way for further investigation.

The bold assumptions that Lenat makes in his new strategy confirms the tendency that Smith observed. As described in the previous section, Lenat assumes away five of the seven elements of shared context for his representational scheme. His model ignores the significance of previous sentences in the transmission of information and the impact of the current situation of the actors. He overlooks the importance of the long term background, roles and achievements of the participants and ignores the significance of the memorable experiences that the participants share. This are bold assumptions because the effort to understand these sort of pragmatic phenomena in AI has given rise to difficulties that challenge the viability of Lenat's classical approach.[313]

A central issue that challenged the viability of Lenat's approach in 1991 and which continues to apply concerns the scope and limits of representation. Bringing together his analysis of important intellectual phenomena, Smith suggested that the full significance of an intentional action can crucially involve non-representational phenomena, as well as representational ones. He expected the analysis of non-representational aspects of an intentional agent to account for salient facts of embodiment, genuine participation in represented subject matters, and the internal manifestation (rather than representation) of intentionally important properties.[314] He concluded:

> 'Lenat and Feigenbaum believe that intelligence can rest entirely on the meaning of representations, without any need for correlated, non-representational experience. On the other hand, they also imagine their system starting to read and distill things on its own. What will happen, however, if the writers tacitly rely on non-representational actions on the part of the reader? The imagined system would not be able to understand what it was reading. For example, there is no way in which their system would ever be able to understand the difference between right and left.'[315]

Lenat has not overcome this difficulty. He does not seem to appreciate that the important pragmatic phenomena which he fails to cater for challenge the viability of the principles, models and methods that he proposes. This is an important weakness in Lenat's new strategy for representing contexts.

Extending Smith's accusation, missing middle realms can be identified in the other

---

[312] Cfr. section 3.2.1.

[313] Cfr. sections 1.2. and 1.3.

[314] Cfr. B. C. Smith. "The owl and the electric encyclopedia", p. 273.

[315] B. C. Smith. "The owl and the electric encyclopedia", p. 275.

simplifications that Lenat makes. To select the dimensions for context-space, Lenat excluded considerations that refer to philosophical, linguistic and other models that are not driven by EH. The result of this is that he continues to ignore the more profound questions that ought to be taken into account in the context of more rigorous and thorough intellectual standards. As mentioned, an important issue that he overlooks is the fact that intelligent agents understand and participate in contexts. The questions that arise here undermine the viability of Lenat's approach. To illustrate this point, Searle's Chinese room argument can be neatly applied. To determine the contexts in which an assertion is valid, a human subject can follow instructions to search for values along 12 independent categories without the least understanding of what those contexts are. There is no reason to suppose that formal definitions of contexts along 12 dimensions can lead to systems that understand contexts.

A further difficulty regarding the choice of dimensions is that the reasons for choosing 12 and not 20 or 200 dimensions are not well articulated. Lenat firmly defends his choice of 12 dimensions but goes on to observe that for most of these dimensions, important sub-dimensions ought to be defined. In some cases, as in the Topic and Culture dimensions, Lenat proposes long lists of sub-dimensions that are incomplete and general. In addition to this some dimensions appear to clump together notions that perhaps require independent treatment. Such is the case with the sixth, seventh and ninth dimensions.[316] Lenat himself admits that his proposal is tentative and that the number and choice of dimensions may change as investigation continues.[317] These realities suggest that the reasons for choosing 12 dimensions lack rigor and completeness. This is a significant weakness in Lenat's proposal.

Together with his selection of 12 dimensions, Lenat makes important simplifications concerning how the dimension values are specified. He claims that most of the information can be captured using inheritance mechanisms, heuristics for automatic update and option lists. A special case that Lenat considers concerns the handling of ambiguity in the conditions in which an assertion is true. Lenat proposes that a handful of persistence distributions can adequately capture ambiguity in these cases. He considers this to be one of the major points of his new proposal. Proceeding in this way, however, Lenat overlooks important aspects of what intelligent agents do. Smith made this observation in his 1991 article explaining that the issues involved are more complex and profound. He wrote:

> 'Both discrete and continuous objects of the sort studied in mathematics (the integers, the real line, and even Gaussian distributions and probability densities) are determinate, in the sense that questions about them have determinate answers. It is unclear, however, in questions about when tea-time ends, or about what adolescence is, or about exactly how many clouds there

---

316 The sixth dimension is Sophistication/Security, the seventh is Topic/Usage and the ninth is Modality/Disposition/Epistemology

317 Cfr. D. B. Lenat, DCS, p. 57.

were when you poked your head out of your tent and said, with complete confidence, "there are lots of clouds today" - it is unclear in such cases whether there are determinate answers at all.'[318]

Smith explains that probability distributions represent determinate phenomena. He suggests that the deeper questions of ambiguity concern phenomena that do not support determinate answers at all. Lenat overlooks these deeper issues in the strategy he proposes to handle ambiguity. This is also an important weakness in his new approach.

To take advantage of the new scheme, new inference mechanisms will be needed. To boost inference, Lenat supposes that sophisticated heuristics and tools will be available for dimensionalizing the contexts of the problems to be solved. He describes the sort of procedures that ought to be supported but fails to go further in illustrating how they can lead to intelligent responses to common sense questions.[319] Neither does he address the difficulties entailed in building the tools he requires. These deficiencies, unless they have already been dealt with at CYC, are important weaknesses in Lenat's new strategy.

Confronted with more rigorous theoretical standards, further weaknesses arise in Lenat's proposal for handling inference. Commenting on Lenat's formal approach to inference in 1991 Smith observed:

'Being able to reason is not just the ability to take the right atomic steps; it means knowing how to think in large - how to argue, how to figure things out, how to think creatively about the world. Traditional logic, of course, does not address these questions. Nor - and this is the important point - is there any a priori reason to believe that that larger inferential demand can be fully met within the confines of logic's peculiar formal and semantic conventions.'[320]

While Lenat has incorporated more sophisticated inference mechanisms since 1991, his approach continues to rely on formal logical techniques and procedures. Smith suggests that phenomena such as knowing how to think in large, how to argue, how to figure things out and how to think creatively are important aspects of how intelligent subjects perform inference using contexts which cannot be reduced to logical mechanisms. He argues that there is no reason to suppose that logic will ever be able to adequately cater for these phenomena. Smith's accusation continues to hold. Lenat leaves aside the more difficult aspects of intelligent behavior that are crucial for appreciating how intelligent subjects reason using contexts. This is a weakness in his new strategy for representing and using context in CYC.

To sum up, in the context of more rigorous and thorough theoretical standards, Lenat's new strategy of context representation has many weaknesses. Smith's central accusation continues to hold. Lenat moves from broad intuition to detailed proposals on

---

[318] B. C. Smith. "The owl and the electric encyclopedia", p. 272.

[319] Cfr. D. B. Lenat, DCS, pp. 69-70.

[320] B. C. Smith. "The owl and the electric encyclopedia", p. 276.

several fronts, assuming away the intermediate conceptual problems that are of central concern in AI. The missing middle realm remains hidden. Lenat strictly adheres to EH and in this context his new strategy receives a much more favorable evaluation. As previously explained, however, the underlying suppositions of EH are mostly conjectural and the theoretical standards that it establishes are weak. Given the availability of profound and useful studies about intelligence and granted the high cost of the project, Smith's insistence that rigorous theoretical standards be employed is well aimed.

4.4. The validity of philosophical evaluation in AI

EH makes it difficult to discuss the strengths and weaknesses of CYC, especially from a philosophical perspective. This is so because it rejects commentaries that are not explicitly based on empirical evidence. In the context of EH progress depends on experiments being able to *falsify* hypotheses and not on philosophical evaluations. A point that ought to be addressed, as a result, is the validity of a philosophical commentary.

In section 2.1.3., it was observed that the epistemology of Karl Popper is the paradigm of the scientific method that Lenat and Feigenbaum uphold for AI. Mariano Artigas has brought to light that the epistemology originally proposed by Popper is more than a simple method, it is an attitude. Artigas stresses the fact that the epistemology of Popper has ethical roots. Progress through falsification and trial and error is part of Popper's idea of rational criticism which, time and again, Popper insists is an attitude and not a theory or doctrine. Falsification and trial and error as a method of investigation is part of a rationality that is rooted in convictions of a philosophical nature. It is a way of proceeding that recognizes the limitations of the human intellect and moderates the dogmatism that leads to confrontation and violence. Falsationism is useful as part of an attitude which says: 'I may be wrong and you may be right, but let us sit together and discuss matters critically, and in the end we may not agree but we will both have learnt something.'[321]

Artigas brings to light that falsification as a scientific method is interesting and useful because the process it prescribes defends against the violent imposition of dogmatic positions. The method of falsification that is widely observed in the scientific community protects and promotes the dignity of the human person. In the light of this, it is clear that falsification as a scientific method strays from its original path if it turns along the path of dogmatism. This is precisely the misfortune that would befall falsification were it to naively insist in embracing hypotheses that go against its moral roots. Such an attitude establishes falsification, no longer as an attitude that others are invited to adopt, but as the only valid method for evaluating ideas. This has the destructive consequence of validating any idea, harmful as it might be to human dignity,

[321] M. Artigas, *Lógica y ética en Karl Popper*, p. 34.

as worthy of serious consideration. If falsification is adopted as a fitting means to healthy dialog and human advancement as originally intended, it needs the light of philosophical evaluation to guide it away from destructive hypotheses and towards those that bolster human dignity.

To summarize, a hypothesis is not worthy of serious consideration simply because it may eliminate errors that have been discovered in experimentation. Aside from its capacity to eliminate errors, a hypotheses gains or loses validity to the extent that it is compatible or not with the philosophical underpinnings that support the rationality of scientific investigation. In this sense, an important weakness of the CYC project is that the attitude of Lenat and his colleagues has been to avoid philosophical discussion and to adhere to falsification as the only method of validation.

Conclusions

The objective of this study was to take time out on behalf of my colleagues in the computer science profession to reflect on the path followed by AI and to examine where we may have strayed from the deeper motivations, desires, and objective limits that ought to configure the way for exciting and fruitful AI research. For such an endeavor, it is necessary to consider concrete projects, seeking out and evaluating the underlying motivations and tendencies. The CYC project gathers together the experience of many years of AI research and is generally considered a forefront project in AI. It is an ideal candidate for the exercise that this work proposed and for this reason this study has centered around the problem of storing common sense in AI. The following paragraphs bring together the principle observations and discoveries that have been made.

1. The term "Artificial Intelligence" is used in many different senses. On the popular front, AI is associated with dazzling exhibits using computers. Electronic devices are indiscriminately described with adjectives connoting intelligence and the media is a welcoming forum for enthusiastic scientists to publicize their futuristic predictions, hopes and desires for AI. In applied AI, the term is normally used in relation to specific concepts or theories that surround the simulation in machines of a particular type of intelligent behavior. Applied AI is often realized in the context of specific military or commercial goals and admits internal techniques that are incompatible with what thinking is about. The terms "human AI" and "Alien AI" are sometimes used to distinguish approaches to artificial intelligence according to whether the objectives and techniques admitted are in accord or not with the demands of a thinking machine. A single project can be considered human AI under certain aspects and alien AI under others. In a scientific context, the term "Artificial Intelligence" is understood as an approach to investigating intelligent behavior using computational models. The models are incorporated in programs and the programs are run to obtain data for evaluation. It is in this paradigm that CYC is best understood.

In the philosophy of AI, the term "Artificial Intelligence" is broadly used, spanning many of the uses previously outlined. It is understood, however, as referring ultimately to the question of the philosophical possibility of a thinking machine. The answer to this question permits a clearer appreciation of the limits and scope of AI and provides orientation for the construction and evaluation of computational models of the mind.

2. Classical AI investigates mental processes by testing logical models of thought. Logical systems are constructed and tested by incorporating them in programs and running the programs to obtain data. It has led to a proliferation of logical systems and their implementation in computers. Some logical techniques such as means-ends analysis, predicate calculus and relational systems have enjoyed a great deal of success. Many others have proven unworkable and there are at present, many experimental systems in progress. CYC adheres to this tradition.

Advocates of classical AI have been severely criticized for ignoring the physical properties and organic operations of the brain in the formulations they propose as models of thought. Another school, connectionist AI, develops its models based on the findings of neuroscience about the physical operations of the brain. While some see in connectionism an alternative to the classical conception of cognition as rule-governed symbol manipulation, others point out that connectionist research is a dead end road if there are no clear means for carrying out the logical processes that characterize thought. Side-stepping the debate, a group of researchers attack both classicism and connectionism for having overlooked many important aspects of human cognition in their research programs. This partial focus, they diagnose, is the cause of the current impasse. These investigators uphold that AI ought to be approached from the ground up, beginning with sensation and having systems that are complete at each stage. Advocates of this view devise and test models by building autonomous mobile robots which interact with their environment. Such investigation is termed “situated robotics” and it is characterized in a special way by its anti-representationalism stance.

3. Those who uphold that computers can be said to think in the same way that we say that humans think, are careful to clarify that not any form of automation can be rightly deemed intelligent behavior. The question at the heart of AI concerns the conditions that have to met in the internal operations of an artifact for there to be intelligent behavior. In AI, three dominant solutions can be identified which correspond to the three dominant trends in AI.

Researchers in classical AI adhere to the Physical Symbol Systems Hypothesis which holds that intelligent behavior can be explained in terms of symbol manipulation. Using the Chinese Room scenario, John Searle has seriously challenged the validity of the Physical Symbol Systems Hypothesis. He argues that there is no reason to suppose that symbol manipulation explains understanding because the English speaker in the Chinese room can move around symbols just as well without understanding what those symbols mean.

Researchers in replacement connectionism propose alternative internal conditions for intelligent behavior. Instead of symbol systems and logical languages, they suggest that intelligence can be explained in terms of changing patterns of connectivity in networks of neurons. Replacement connectionists maintain that a connectionist cognitive architecture would contain sentence-like representations as a phenomenon emerging from the network. At present there is no consensus as to exactly how sentence like structures can emerge from a network without appealing to classical models.

Furthermore Searle points out that connectionist systems, while closer to the brain, continue to be collections of formal procedures. He argues that in principle a person can internalize the formal structure of a connectionist network, and do all the neuron firing and adjusting in his or her imagination, exactly mimicking the real system. The person may produce correct answers in Chinese without understanding any Chinese at all. These difficulties have greatly jeopardized the connectionist endeavor.

Researchers in anti-representationalist robotics propose a third alternative regarding the internal conditions for intelligent behavior. Investigators in this field emphasize that the modules ought to be directly behavior-producing. Higher-level functions such as perception, planning, modeling and learning emerge from the interaction of these lower level modules. While Searle commends anti-representationalist robotics for conceding that cognition is not solely a matter of formal symbol manipulation, he argues that the model is insufficient. If the computer inside the robot is replaced with a person, the person is capable of realizing the behavior exhibited by the robot without understanding anything at all.

4. The Turing test in some form or other is the point of reference for those who defend that computers can be said to think in the same manner that we say humans think. The Test affirms that an artifact that exhibits what is considered to be intelligent behavior can be considered truly intelligent. The Test itself proposes a concrete scenario in which intelligent behavior is characterized by the ability to respond intelligently to a wide variety of questions. Searle's scenario of the Chinese Room, however, illustrates in a graphic way that the Turing test is simply insufficient. Some philosophers suggest that the issue can be dissolved by changing the way we use language to speak about artifacts. Dennett proposes a strategy which he calls the *intentional strategy* or adopting the *intentional stance*. According to this strategy, a system is intentional when its behavior is predictable and explainable in terms of beliefs and desires we attribute to it together with the rationality required to figure out what it ought to do in the light of those beliefs and desires. While his writings have enjoyed popular appeal, Dennett's strategy leaves aside an important aspect of beliefs. Beliefs are not merely predictive strategies; they can be objectively judged true or mistaken, founded or unfounded, reasonable or fictitious. Furthermore, the Chinese room scenario articulated by John Searle illustrates that it is an error to attribute beliefs and desires to artifacts.

Choosing a different linguistic approach, Copeland argues that the debate over whether an artifact can be a thinking thing can be resolved by a decision on the part of the linguistic community as to what definition of thought best serves our purposes in the modern world. He argues that the purposes for which we use the concept of a thinking thing in the new age of computer technology are best served if we indeed decide to count an appropriately programmed computer as a thing that thinks. Such a move, however, would impoverish the way we speak about intelligence as it excludes a reference to the peculiar internal activity in human thought which the Chinese Room scenario brings to light and which the pragmatic tradition has shown to be an essential part of intelligence.

5. Programs in the early days of AI relied heavily on compiled knowledge and formal methods for problems solving. They were highly competent is certain limited domains, but entirely useless in others. Deeper issues of general intelligence were largely ignored. AI experts generally agree that one of the limiting factors has been a lack of knowledge. An important lesson learned over the years is that general intelligence cannot be separated from general knowledge. Knowledge is needed to communicate, knowledge is needed to reason, knowledge is needed to learn. Lenat's CYC project is a bold effort that takes up the challenge to test models of general intelligence that are knowledge based.

6. Lenat and Feigenbaum summarize what they consider to be the appropriate methodology for the science of Artificial Intelligence in their Empirical Inquiry Hypothesis (EH). EH adapts the methodology of empirical investigation to AI. It establishes that AI investigation is driven by a process of falsification. The hypotheses are embodied in programs and the programs are run to obtain data. This data is evaluated according to the sophistication of the surprises and the level of performance in a wide variety of domains. During testing important limitations are observed. These limitations suggest ways that the original hypotheses can be improved. EH stipulates that the new hypotheses that are proposed *must* address the shortcomings that were encountered while safeguarding prior achievements. CYC theory is a product of this dynamic and it is in this context that it ought to be understood.

7. Linked to EH, the Knowledge Principle, the Explicit Knowledge Principle and the Breath Hypothesis are the principle postulates of CYC theory. The Knowledge Principle affirms that a system exhibits intelligent understanding and action at a high level of competence primarily because of the *knowledge* that it can bring to bear: the concepts, facts, representations, methods, models, metaphors, and heuristics about its domain of endeavor. The Explicit Knowledge Principle asserts that for knowledge to be useful, it needs to be explicitly defined. The essential characteristics of the knowledge items ought to be specified in a consistent way and it ought to be clearly indicated how the knowledge items relate to each other. Finally, the Breath Hypothesis is a mandate to build systems that incorporate two fundamental strategies for dealing with novelty. On one hand, the system must have the capacity to fall back on increasingly general knowledge and secondly, the system ought to possess the ability to reason by analogizing to specific knowledge from far-flung domains.

8. The most important lesson learnt over the years in building CYC is the need to incorporate contexts. Lenat and his team initially envisioned a flat context-free knowledge base of common sense knowledge. As more items were added it became more urgent to avoid the useless repetition of assumptions for assertions that apply in similar situations. Furthermore, lacking a consistent way to link groups of assertions and assumptions it became more difficult to ensure consistency. A new strategy was needed which would resolve these shortcomings. In response to this challenge, Guha developed models and techniques for grouping together assertions and assumptions into contexts or microtheories. By means of powerful copy/edit functions, new contexts

could be easily created by "lifting" assertions and assumptions from other contexts into the new context. Using this strategy, the CYC KB was carved up into hundreds of contexts.

While microtheory technology promised to overcome the early difficulties, important shortcomings arose which undermined and offset its usefulness. Lifting turned out to be a bothersome source of logical inconsistencies. Furthermore, as more contexts were created it became difficult and time consuming for the knowledge enterer to select the most adequate context each time he/she updated the KB with new information. Finally, Lenat and his team were faced with the difficult challenge of maintaining a good intermediate granularity for the number and choice of contexts. Though microtheory technology provided ways to overcome early difficulties, the shortcoming encountered during its implementation suggested the need for new theories and techniques, starting again the EH cycle.

Lenat's new strategy addresses the shortcomings of microtheory technology. He proposes that context can be adequately defined by specifying values along 12 independent dimensions. Proceeding in this way, Lenat upholds that the difficulties encountered in the previous strategy can be overcome.

9. Lenat strictly adheres to EH. The benefits that come from this method are similar to those that derive from the empirical method of research in general. The process of falsification that the empirical or scientific method entails guarantees a measure of progress. The experiments are well defined and repeatable. They provide a context which permits the scientific community to objectively determine whether a new hypotheses represents an advance or not over older theories.

10. While EH may strengthen the relevance of CYC in terms of objectivity and progress, several weaknesses arise when it's method is carefully confronted with the general paradigm of scientific investigation. AI, pretends to investigate real or natural intelligence. The methodology of the empirical sciences in general demands of AI a reference to natural intelligence. The method of AI adopted in CYC has the peculiar characteristic that the reference is established through the mediation of digital machines. In their exposition of their method, Lenat and Feigenbaum do not state explicitly how the reference to natural intelligence is established. They tacitly suppose that an adequate reference is provided for by means of the Physical Symbol Systems Hypothesis. They refer to the Physical Symbol Systems Hypothesis as 'an article of faith'. In consequence, the reference established through it is weak and uncertain. Researchers in replacement connectionism consider the hypothesis untenable because it ignores the physical properties and organic operations of the brain. Researchers in anti-representationalist robotics consider the Physical Symbol Systems Hypothesis presumptuous and misleading because it overlooks many important aspects of human cognition such as situatedness, embodiment and emergence.

11. Linked to their reliance on the Physical Symbol Systems Hypothesis, empirical inquiry demanded of the authors a way of evaluating the extent to which a computer has

been appropriately programmed. The authors suppose that a computer system that demonstrates what is considered to be intelligent conduct (interesting surprises and performance in a wide selection of domains) can be considered, to some degree, really intelligent. Lenat and Feigenbaum suppose an immensely simplified version of the Turing Test, more in accord with the models and techniques at their disposition. Their criteria are vague and imprecise, a far cry from what Turing had in mind.

12. The underlying suppositions of EH (points 9 and 10) gives rise to low standards of rigor and thoroughness for the hypotheses that are proposed in CYC. Difficult issues can be assumed away without violating theoretical standards in the context of EH. An important weakness in CYC theory is that it fails to take into account important phenomena that have been studied and investigated for many years. Similarly, in the context of more rigorous and thorough theoretical standards, Lenat's new strategy for context representation assumes away many difficult though important issues. Smith's central accusation continues to hold. Lenat moves from broad intuition to detailed proposals on several fronts, side stepping the intermediate conceptual problems that are of central concern in AI.

13. The methodology that CYC follows rejects commentaries that are not explicitly based on empirical evidence. Progress depends on experiments being able to *falsify* hypotheses and not on philosophical evaluations. Mariano Artigas has shown that falsification and trial and error as a method of investigation is rooted in the epistemology of Karl Popper. It is a way of proceeding that recognizes the limitations of the human intellect and moderates the dogmatism that leads to confrontation and violence. As a result of this, a hypothesis is not worthy of serious consideration, simply because it may eliminate errors that have been discovered in experimentation. Aside from its capacity to eliminate errors, a hypotheses gains or loses validity to the extent that it is compatible or not with the philosophical underpinnings that support the rationality of scientific investigation. An important weakness of the CYC project is that the general attitude is to avoid philosophical discussion and to adhere falsification and trial and error as the only method of validation.

A consequence of Lenat's attitude towards philosophy is that he will never confront the deeper questions of AI investigation that challenge the underlying principles of the method that he adopts. EH commits Lenat to rely on the Physical Symbol Systems Hypothesis and to suppose a weak version of the Turing Test. Lenat's strict adherence to EH allows and drives him to assume away difficult questions that challenge the viability of his principle postulates. The result is that there will always be a missing middle realm in the theories and models that Lenat proposes. This is an unfortunate fate because, as Smith points out, a great deal of work has already been done to better understand the deeper and more difficult issues involved. This work, based on intellectual rigor and thoroughness and a firm trust in human reason, permits a clearer appreciation of the limits and scope of AI and provides orientation for the construction and evaluation of computational models of the mind. Lenat refuses to avail himself of the guidance and support that comes with sound philosophical study. His position is

contradictory because the empirical investigation that he advocates supposes a philosophical commitment which ought to be evaluated in the context of AI. Lenat violently imposes method over intellectual rigor and thoroughness. This is a deep underlying tendency in CYC and a constant danger for AI research. The result is that CYC is vulnerable to destructive hypothesis that are imaginary and harmful. Such hypotheses lead to a dead end and to the frustration of having wasted a great deal of time, effort and resources.

Bibliography

Works by Douglas Lenat and his associates

Blair, P., R. V. Guha and W. Pratt, 1992, "Microtheories: An Ontological Engineer's Guide", *MCC Technical Report* CYC-050-92, Austin, TX: Microelectronics and Computer Technology Corporation.

Guha, R. V., 1990, "Contexts in CYC", *MCC Technical Report* CYC-129-90, Austin, TX: Microelectronics and Computer Technology Corporation.

Guha, R. V., 1990, "The Representation of Defaults in CYC", *MCC Technical Report* CYC-083-90, Austin, TX: Microelectronics and Computer Technology Corporation.

Guha, R. V., 1993, "Context Dependence of Representations in CYC", *MCC Technical Report* CYC-066-93, Austin, TX: Microelectronics and Computer Technology Corporation.

Guha, R. V., and A. Levy, 1990, "A Relevance-Based Meta Level", *MCC Technical Report* CYC-450-90, Austin, TX: Microelectronics and Computer Technology Corporation.

Heyes-Roth, F., D. A. Waterman and D. Lenat, eds., 1983, *Building Expert Systems,* Reading, MA: Addison-Wesley.

Lenat, D. B., 1998, "The Dimensions of Context Space", Austin, TX: Cycorp.

Lenat, D. B., 1997, "From *2001* to 2001: Common Sense and the Mind of HAL" in D. G. Stork, ed., 1997, *Hal's Legacy: 2001's Computer as Dream and Reality*, Cambridge, MA: MIT Press, pp. 195-220.

Lenat, D. B. and E. A. Feigenbaum, 1991, "On the Thresholds of Knowledge", *Artificial Intelligence* 47: 185-230.

Lenat, D. B. and E. A. Feigenbaum, 1991, "Reply to Brian Smith", *Artificial Intelligence* 47: 230-250.

Lenat, D. B. and J. S. Brown, 1984, "Why AM and Eurisko Appear to Work", *Artificial Intelligence* 23: 269-294.

Lenat, D. B. and R. V. Guha, 1990, *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project*, Reading, MA: Addison-Wesley.

Lenat, D. B. and R. V. Guha, 1990, "CYC: A Mid-Term Report", *AI Magazine* Fall: 32-59.

Lenat, D. B. and R. V. Guha, 1994, "Enabling Agents to Work Together", *Communications of the ACM* 37: 127-142.

Lenat, D. B., M. Prakash and M. Shepherd, 1986, "CYC: Using Common Sense Knowledge to overcome Brittleness and Knowledge acquisition bottlenecks", *AI Magazine* Summer: 65-85.

Lenat, D. B., R. V. Guha, K. Pittman, D. Pratt and M. Shepherd, 1990, "CYC: Towards programs with Common Sense", *Communications of the ACM* 33: 30-49.

General bibliography

Artigas, M., 1994, *El desafío de la racionalidad*, Pamplona: Eunsa.

Artigas, M., 1994, *La filosofía de la ciencia experimental*, Pamplona: Eunsa.

Artigas, M., 1999,*Lógica y ética en Karl Popper*, Pamplona: Eunsa.

Austin J. L., 1962, *Philosophical Papers*, Oxford: Clarendon Press.

Austin J. L., 1975, *How to Do Things With Words*, Cambridge, MA: Harvard University Press.

Austin, J. L., 1964, *Sense and Sensabilia*, New York: Oxford University Press.

Barnett, J. and K. Knight, 1990, "Knowledge and Natural Language Processing", *Communications of the ACM* 33: 50-71.

Bickhard M. H. and L. Terveen, 1995, *Foundational Issues in Artificial Intelligence*, Amsterdam: Elsevier.

Bloomfield, B., ed., 1987, *The Question of Artificial Intelligence: Philosophical and Sociological Perspectives*, London: Helm.

Boden, M., ed., *The Philosophy of Artificial Intelligence*, Oxford: Oxford University Press, 1990.

Boden, M., 1998, "Artificial Intelligence" in *Routledge Encyclopedia of Philosophy* vol. 1, London: Routledge, pp. 485-492.

Born, R., ed., 1988, *Artificial Intelligence: The Case Against*, London: Routledge.

Brachman, R. and H. Levesque, eds., 1985, *Readings in Knowledge Representation*, Los Altos, CA: Morgan Kaufmann.

Broadbent, D., 1993, *The Simulation of Human Intelligence*, Oxford: Blackwell.

Brooks, R., 1991, "Intelligence without Representation" in J. Haugland, ed., *Mind Design II*, Cambridge, MA: MIT Press, 1997, pp. 395-420.

Burkholder, L., 1992, *Philosophy and the Computer*, Boulder, CO: Westview Press.

Button, G., ed., 1995, *Computers, Mind and Conduct*, Cambridge, MA: Polity Press.

Campbell, M., 1997, "How HAL Plays Chess" in D. G. Stork, ed., 1997, *Hal's Legacy: 2001's Computer as Dream and Reality*, Cambridge, MA: MIT Press, pp. 75-100.

Charniak, E. and D. McDermott, 1985, *Introduction to Artificial Intelligence*, Reading, MA: Addison-Wesley.

Churchland, P., 1988, *Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind*, Cambridge, MA: MIT Press.

Clark, A., 1998, *Being There: Putting Brain, Body and World Together Again*, Cambridge, MA: MIT Press.

Collins, H. M., 1993, *Artificial Experts: Social Knowledge and Intelligent Machines*, Cambridge, MA: MIT Press.

Copeland, J., 1993, *Artificial Intelligence: A Philosophical Introduction*, Oxford: Blackwell.

Dennett, D., 1981, "True Believers: The Intentional Strategy and Why it Works" in J. Haugland, ed., *Mind Design II*, Cambridge, MA: MIT Press, 1997, pp. 57-80.

Dreyfus, H., 1992, *What Computers Still Can't Do, A Critique of Artificial Reason*, Cambridge MA: The MIT Press.

Ernst, G. W. and A. Newell, 1969, *GPS: A Case Study in Generality and Problem Solving*, New York: Academic Press.

Fetzer, J. H., 1990, *Artificial Intelligence: It's Scope and Limits*, Dordrecht: Kluwer.

Fodor, J. A., 1975, *The Language of Thought*, New York: Thomas Crowell.

Fodor, J. and Z. Pylyshyn, 1988, "Connectionism and Cognitive Architecture: A Critical Analysis", in J. Haugland, ed., *Mind Design II*, Cambridge, MA: MIT Press, 1997, pp. 309-350.

Garnham, A., 1988, *Artificial Intelligence: An Introduction*, London: Routledge.

Genova, G., 1997, *Charles S. Peirce: La lógica del descubrimiento*, Pamplona: Servicio de Publicaciones de la Universidad de Navarra.

Ginsberg, M., ed., 1987, *Readings in Nonmonotonic Reasoning*, San Mateo, CA: Morgan Kaufmann.

Goldkind, S., 1990, *Machines and Intelligence: A Critique of Arguments Against the Possibility of Artificial Intelligence*, New York: Greenwood Press.

Grossman, L., 1998, "Get Smart: How Intelligent Can Artificial Intelligence Be? Lev Grossman Talks To Doug Lenat To Find Out", *Time-Digital Magazine* (October 1998).

Hanks, S. and D. McDermott, 1986, "Default Reasoning, Nonmonotonic Logic, and the Frame Problem", *Proceedings of the Fifth National Conference on Artificial Intelligence*: 328-333.

Hodgson, J. P. E., 1991, *Knowledge Representation and Language in AI*, New York: Ellis Horwood.

Horgan, T. and J. Tienson, 1995, "Connectionism and the Commitments of Folk Psychology", *Philosophical Perspectives* 9: 127-152, (1995).

Houser, N. and C. Kloesel, eds., 1992, *The Essential Peirce: Selected Philosophical Writings*, vol. 1-2, Bloomington, IN: Indiana University Press.

Jaki, S., 1989, *Brain, Mind and Computers*, Washington: Gateway.

Ketner, K., ed., 1992, *Reasoning and Logic of Things: Charles S. Peirce*, Cambridge, MA: Harvard University Press.

Kuhn, T. S., 1970, *The Structure of Scientific Revolutions*, 2nd ed., Chicago, IL: University of Chicago Press.

Kulas, J., J. Fetzer and T. Rankin, eds., 1990, *Philosophy, Language and Artificial Intelligence: Resources for Processing Natural Language*, Dordrecht: Kluwer.

Kurzweil, R., 1990, *Intelligent Machines*, Cambridge, MA: MIT Press.

Kurzweil, R., 1997, "When Will HAL Understand What We Are Saying? Computer Speech Recognition and Understanding" in D. G. Stork, ed., 1997, *Hal's Legacy: 2001's Computer as Dream and Reality*, Cambridge, MA: MIT Press, pp. 131-170.

Lakoff, G. and M. Johnson, 1980, *Metaphors We Live By*, Chicago, IL: University of Chicago Press.

MacNamara, J. and G. E. Reyes, eds., 1994, *The Logical Foundations of Cognition*, Oxford: Oxford University Press.

McCarthy, J. and P. J. Hayes, 1987, "Some Philosophical Problems from the Standpoint of AI" in M. Ginsberg, ed., *Readings in Nonmonotonic Reasoning*, 523-531, San Mateo, CA: Morgan Kaufmann.

McCarthy, J., 1986, "Applications of Circumscription to Formalizing Common-Sense Knowledge", *Artificial Intelligence* 28: 89-116.

Minsky, M. and S. Papert, 1969, *Perceptrons: An Introduction to Computational Geometry*, Cambridge, MA: MIT Press.

Monk, R., 1991, *Ludwig Wittgenstein: The Duty of Genius*, London: Vintage Press.

Moore, R. C., 1986, "The Role of Logic in Knowledge Representation and Common Sense Reasoning" in H. Levesque and R. Brachman, eds., *Readings in Knowledge Representation*, San Mateo, CA: Morgan Kaufmann, pp. 335-343.

Nilsson, N., 1980, *Principles of Artificial Intelligence*, Palo Alto, CA: Tioga.

Nubiola, J., 1994, "C. S. Peirce: pragmatismo y logicismo", *Philosophica* 17, 1994, pp. 209-216.

Nubiola, J., 1994, *La renovación pragmatista de la filosofía analítica. Una introdución a la filosofía contemporánea del lenguaje*, Pamplona: Eunsa.

Popper, K., 1959, *The Logic of Scientific Discovery*, London: Hutchinson.

Ramsey, W., S. Stich and J. Garon, 1990, "Connectionism, Eliminativism, and the Future of Folk Psychology" in J. Haugland, ed., *Mind Design II*, Cambridge, MA: MIT Press, 1997, pp. 351-376.

Rapaport, W. J., 1988, "Syntactic Semantics: Foundations of Computational Natural Language Understanding" in J. Fetzer, ed., *Aspects of Artificial Intelligence,* Dordrecht: Kluwer, pp. 81-131.

Rich, E. and K. Knight, 1991, *Artificial Intelligence*, New York: McGraw-Hill.

Rosenfeld, A., 1997, "Eyes for Computers: How HAL Could See" in D. G. Stork, ed., 1997, *Hal's Legacy: 2001's Computer as Dream and Reality*, Cambridge, MA: MIT Press, pp. 211-236.

Rumelhart, D. E., 1989, "The Architecture of Mind: A Connectionist Approach" in J. Haugland, ed., *Mind Design II*, Cambridge, MA: MIT Press, 1997, pp. 205-232.

Schank, R., 1997, "How Could HAL Use Language" in D. G. Stork, ed., 1997, *Hal's Legacy: 2001's Computer as Dream and Reality*, Cambridge, MA: MIT Press, pp. 171-190.

Searle, J., 1980, "Minds, Brains, and Programs" in R. Born, ed., 1988, *Artificial Intelligence: The Case Against*, London: Routledge, pp. 18-40.

Sejnowski, J. and C. Rosenberg, 1987, "Parallel Networks That Learn to Pronounce English Text", *Complex Systems* I (1987), pp. 145-168.

Shasha, Dennis and Cathy Lazere, *Out Of Their Minds: The Lives and Discoveries of 15 Great Computer Scientists*, New York: Copernicus, 1995.

Shortliffe, E. H., 1976, *Computer-Based Medical Consultations: MYCIN*, New York: Elsevier.

Smith, B. C., 1991, "The Owl and the Electric Encyclopedia", *Artificial Intelligence* 47, 1991, pp. 251-288.

Smolensky, P., 1989, "Connectionist Modeling: Neural Computation/Mental Connections" in J. Haugland, ed., *Mind Design II*, Cambridge, MA: MIT Press, 1997, pp. 233-250.

Stork, D. G., 1997, "Scientist on the Set: An Interview with Marvin Minsky" in D. G. Stork, ed., 1997, *Hal's Legacy: 2001's Computer as Dream and Reality*, Cambridge, MA: MIT Press, pp. 16-33.

Stork, D. G., ed., 1997, *Hal's Legacy: 2001's Computer as Dream and Reality*, Cambridge, MA: MIT Press.

Turing, A. M., 1950, "Computing Machinery and Intelligence", *Mind* 59: 33-60.

Ullman, J. D., 1989, *Principles of Database and Knowledge-Base Systems*, Vol. 1, Rockville, MD: Computer Science Press.

Weiner, P., ed., 1966, *Charles S. Peirce: Selected Writings. Values in a Universe of Chance*, New York: Dover.

Weizenbaum, J., 1966, "ELIZA - a Computer Program for the Study of Natural Language Communication Between Man and Machine", *Communications of the ACM* 9: 36-45.

Winograd, T. A., 1972, *Understanding Natural Language*, New York: Academic Press.

Winograd, T. and F. Flores, 1986, *Understanding Computers and Cognition*, Norwood, NJ: Ablex.

Wittgenstein, L., 1972, *On Certainty*, New York: Harper Torchbooks.

Wittgenstein, L., 1997, *Philosophical Investigations*, Oxford: Blackwell.

Wittgenstein, L., 1997, *Tractatus Logico-Philosophicus*, London: Routledge.

Wolff, J. G., 1991, *Towards a Theory of Cognition and Computing*, New York: Ellis Horwood.