MAKING MACHINES IN OUR IMAGE: THE RHETORIC OF ARTIFICIAL INTELLIGENCE

A Dissertation

by

ROY JOSEPH

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2001

Major Subject: Speech Communication

UMI Number: 3033810

UMI

UMI Microform 3033810

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

MAKING MACHINES IN OUR IMAGE: THE RHETORIC OF ARTIFICIAL INTELLIGENCE

A Dissertation

by

ROY JOSEPH

Submitted to Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Marshall Scott Poole
(Co-Chair of Committee)

Charles Conrad
(Co-Chair of Committee)

Vanessa Beasley
(Member)

Colin Allen
(Member)

Richard Street
(Department Head)

December 2001

Major Subject: Speech Communication

ABSTRACT

Making Machines in Our Image: The Rhetoric of Artificial Intelligence

(December 2001)

Roy Joseph, M.A., University of Nevada-Las Vegas; M. Phil., Pondicherry University, India; M.A., Pondicherry University, India; B.A., Pondicherry University, India

Co-Chairs of Advisory Committee: Dr. Marshall Scott Poole
Dr. Charles Conrad

With the emergence of Artificial Intelligence, metaphysical accounts of mind are increasingly replaced by computational psychology based on the presupposition that humans are machines in spite of differences in 'hardware.' The alliance of computational psychology with the neurosciences undergirded by a strong materialism facilitated the stage for the acceptance of Artificial Intelligence as a model of mind. Artificial Intelligence (henceforth referred to as AI) claims to offer a resolution to the question 'what is mind?' by asserting that an efficacious method of studying minds lies in the act of building them. Most discourse of philosophy of mind on AI centers on a nonconstructivist view of language.

A rhetoric of mind, on the other hand, tells us about the creative nature of our linguistic frames; it tells us about the role metaphors play in generating conceptual vocabularies and theoretical frameworks; it tells us about the interrelationships between concepts and how these concepts are co-opted into frameworks and how these frameworks become pivotal in our understanding of mind, and finally human nature.

Having said so, the basic thesis of this discussion is that the discourse of mind implicit in AI is also rhetorically constructed as seen through: a) the emergent rhetorical situation of scientific materialism and b) the role of generative metaphors in creating concepts and theoretical models of cognition. How these metaphors developed and play out in the discourse of mind will be the key focus of this study.

DEDICATION

This thesis is dedicated to my parents, siblings, nephew, and to the memory of my grandmother.

ACKNOWLEDGMENTS

I honestly want to begin by thanking my elementary school teachers who taught me how to spell. After typing the first sentence, I just realized that there have been myriads of people who played both constructive and 'not-so-constructive' roles in my life. However, for the present moment, a prolonged documentation of the 'who's who' in my life is not necessary, since all my friends and well-wishers will always occupy a special place in my heart. Therefore, I shall begin by acknowledging the faculty and my classmates in the Department of Speech Communication, Texas A&M University (with special regard for my friends Buddy and Dave who are on their way to becoming top-notch rhetorical scholars).

More importantly, words cannot express the gratitude I feel towards two people who have played a special role as my mentors – Dr. Scott Poole and Dr. Charles Conrad. I want to thank Dr. Poole not only for his time and patience, but also for his invaluable contribution towards my intellectual development. He opened up my world to new ways of thinking, fresh insights, tomes of scholarship among other things. Besides his extraordinary erudition, I have also been specially touched by his humility and willingness to mentor me. Thank you Dr. Poole for your suggestions and assistance. Dr. Charles Conrad also occupies a special place with regard to my intellectual development. Besides his heart-warming sense of humor, his mentorship both as graduate advisor and as my co-chair has been par excellence. Thank you for walking me through this program, making me feel welcome and also for your 'edits', not to mention your entertaining editorializing in the margins of my research papers that provided very good social commentary about the interesting times we live in.

My gratitude extends to both Dr. Beasley and Dr. Allen as well, for their constructive criticism and encouragement throughout the process.

Last but not least, I want to thank my Savior for not giving up on me.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

G. K. Chesterton once remarked "the function of the imagination is not to make strange things settled, so much as to make settled things strange; not so much to make wonders facts as to make facts wonders."[1] In the spirit of Chesterton's maxim, let us invite readers to revisit a question that has been raised since classical antiquity – 'what is the nature of mind? It is a truism to say that the entire gamut of human behavior involves the mind. This question rarely crosses our consciousness as we shuffle through the predictable monotony of every day living, yet, it is central to unpacking fundamental issues pertaining to the nature of human consciousness, mind-body interaction, 'other minds,' and perhaps most importantly the ontological question of human nature itself. This question has been framed differently by opposing philosophical traditions.

Speculative metaphysics characterizes the mind as a distinct, non-physical 'space,' a notion that has solidified over time by the complex sedimentation of mentalistic vocabularies – Platonic idealism, Boethius' rationalism, Berkeley's mental monism, Hegel's universal Spirit only to name a few, have perpetuated the view pertaining to the superiority of the mind over the body. Physicalism with its renewed commitment to a revitalized materialism, treats the mind as only a physical entity. This notion was given currency a notion by the natural philosophers of ancient Greece and has been refined with greater perspicuity by computational psychology and the

---

The journal model for this dissertation is *Rhetoric & Public Affairs*.

neurosciences with their fine eye for detail.[2] While speculative metaphysics inadvertently promotes a species-centered humanism, a materialist philosophy of mind on the contrary extends the scope of intelligence to non-biological or non-human phenomena as well, concatenating in a symbolic fashion the neural circuitry of biological brains with digital computers. With the emergence of Artificial Intelligence, metaphysical accounts of mind are increasingly replaced by computational psychology based on the presupposition that humans are machines in spite of differences in 'hardware.' The alliance of computational psychology with the neurosciences undergirded by a strong materialism facilitated the stage for the acceptance of Artificial Intelligence as a model of mind.

Artificial Intelligence claims to offer a resolution to the question 'what is mind?' by asserting that an efficacious method of studying minds lies in the act of building them. The enterprise of mind design or building artificial minds yield productive models of cognition that shed light on how the mind works. And indeed, 'the map is the territory' is the story of intelligence as told by cognitive science seasoned with finely textured analyses and an ever-growing body of literature possessing noteworthy analytical depth and acute philosophical rigor. As contributions from cognitive science continue to grow, what can perhaps be added to the conversation table is a rhetorical perspective. In other words, while there are many scientific findings in the AI area, what has been less considered is the rhetorical grounding of AI, the means by which its rhetoric creates a unique view of mind.

The term 'rhetoric of mind' has never been used before

The juxtaposition of epistemology with rhetoric is fully intentional, in that, this dissertation will argue that our choice of language plays an instrumental role in the manner in which we see the world, and that science, with its professed impersonality, objectivity and deep-seated suspicion of the human subject also uses linguistic lenses. The lenses themselves are reflections of intentional, discursive practices within the scientific community whereby agents select generative metaphors that play a creative role in shaping the discourse. How this is the case can be seen through close scrutiny to the production of scientific discourse.

Why a rhetoric of mind when perhaps a philosophy of mind would suffice? Of course, the existence of the former is impossible without the latter. Most discourse of philosophy of mind on AI centers on a nonconstructivist or scientific realist view of language.[3] A rhetoric of mind, on the other hand, tells us about the constructive nature of our linguistic frames, it tells us about the role metaphors play in generating conceptual vocabularies and theoretical frameworks, it tells us about the interrelationships between concepts and how these concepts are co-opted into frameworks and how these frameworks become pivotal in our understanding of mind, and finally human nature.

Having said so, the basic thesis of this discussion is that the discourse of mind implicit in AI is also rhetorically constructed as seen through: a) the emergent rhetorical situation of scientific materialism and b) the role of generative metaphors in creating concepts and theoretical models of cognition. A rhetorical situation in the words of Lloyd F. Bitzer is simply this: "When I ask, what is a rhetorical situation?, I want to know the nature of the contexts in which speakers or writers create rhetorical discourse."[4] The contexts help us understand how the discourse came about, and the generative metaphors help us understand how the discourse is generated and sustained. The selection of a particular metaphor may perhaps be tacit, however a metaphor once used commonly within the community becomes a useful resource to provide direction to the discourse in generating a conceptual vocabulary.

In order to trace the rhetorical underpinnings of AI, the dissertation will first undertake an analysis of the rhetorical situation facilitating the emergence of AI, with specific reference to Alan Turing's postulations on machine intelligence, symbolic and connectionist AI respectively. Following this it will discuss the rhetorical construction of metaphoric models of 'thought' in each camp using Kenneth Burke's conception of a terministic screen. The goal is not only to describe what the rhetoric *is*, namely the use of metaphors, but also to discuss how the rhetoric plays out in the generation of frameworks when metaphors are used. The metaphors play out in the development of theoretical concepts that are used as cognitive indices of human mental processes. The metaphoric model serves as the epistemological grounds on which the tacit assumptions of cognition presupposed by each camp about the mind are supported and developed.

Preview of Arguments

Scientific materialism emerged as a sophisticated response to an intellectual exigency, namely the need to provide an explanation to a wide range of conceivable phenomena, by foregrounding 'matter' as an ontological necessity to make any meaningful scientific description. Although scientific materialism is a philosophical movement with robust variations, a unified theme that binds scientific materialists (especially exclusive materialists) is the notion that considering the mind as a physical entity alone is the only answer to questions pertaining to the nature of mind. Materialist philosophies of mind respond to the question 'what is mind?' by stating that it is an emergent property of matter and that Artificial Intelligence clearly demonstrates how this is so. Such an act of demonstration falls back on models that are capable of generating a descriptive and conceptual vocabulary. The models of mind found in AI literature are metaphoric representations of cognition, whereby the explanatory appeal of a particular metaphor warrants its selection over another.

Given the substantive length of this project and conceptual density of the technical literature, a brief preview of the arguments germane is in order. If AI is rhetorically constructed as the thesis claims, then it is important to address a pertinent question 'how can AI be seen as rhetoric?' Therefore, this discussion should discuss what the rhetoric is, and also talk about how the rhetoric plays out through the respective terministic screens emergent from Turing's discourse, symbolic AI and connectionist AI.

In that light, Chapter 2 will attempt to provide a general theoretical framework by examining the use of metaphors as terministic screens and the dynamic skills of practical argumentation based on analogical reasoning. The purpose of this chapter is to situate the rhetoric of Artificial Intelligence within the larger framework of the rhetoric of science. Chapter 3 sees anthropomorphizing as a deep-seated motivation for undertaking any type of epistemological inquiry. Just as metaphors are the terministic screens elucidating the rhetorical dimensions of AI, anthropomorphizing (Chapter 3) is its rhetorical motivation. However, it must be mentioned that the sense in which anthropomorphizing is employed is different from the traditional usage. The conventional use sees anthropomorphizing merely as a projection of human qualia onto animals or artifacts. However, the usage in itself tells us little about the motivation embedded in such a rhetorical gesture. The sense in which anthropomorphizing is used here simply means that we use a particular lens or framework that makes most sense to us, in order to understand something that we consider to be similar – ultimately, it boils down to the motivational question that we try to make sense of the world in a language that we create, and therefore the idea of apprehending reality comes back to the language that we have created in order to understand it.

How does anthropomorphizing fit in with the thesis that the discourse of mind in AI is rhetorically constructed? First of all, the very assertion that AI is rhetorical may seem incorrect to honest thinkers who are dedicated to a fact-based picture of the world. The rhetorical turn is based on the claim that a seemingly simple fact, pristine and untainted in its locality has to be apprehended linguistically to warrant its very recognition as a fact. The recognition of the need for some type of linguistic vehicle is the starting point for recognizing the value of a rhetorical understanding of science. The view that we do indeed require a language presupposes the involvements of agents who use this language, and the view that there are agents who employ this language, posits agents as carriers of assumptions. The presence of these assumptions imply that frameworks (based on intermeshing these assumptions) are brought in, and the bringing in of a framework implies the bringing in of prior understandings (although the understanding may be subjected to subsequent revisions), and the bringing in of a prior understanding implies the bringing in of an attempt to understand the world in terms of what is linguistically readily available.

A series of such gestures involved in the human, epistemological enterprise are ultimately mediated persuasively regardless of whether one is a scientific realist or a constructivist. At this point, one may ask why is persuasion involved? This can be illustrated with an example – it was necessary to persuade both scientific audiences and the laity about the scientific probability of evolution, if evolutionary theory were to used as a framework or cornerstone to approach other related biological issues. By the same token, the machine metaphor is indispensable and its usefulness has to be communicated persuasively if computational psychology is to gain credence beyond its circle of adherents. When rhetoric points in the direction of discursive or language moves employed by scientific rhetors, anthropomorphizing points in the direction of agential motivation behind the rhetorical moves. Michael Polanyi best explains the idea of agential motivation expressed through 'purpose' and 'commitment' that are involved in understanding:

> But the context of purpose and commitment, as found inherent in the personal contribution of the knower to his knowledge, yet lacks dynamic character. The pouring out of ourselves into the particulars given by experience so as to make sense of them for some purpose or in some coherent context, is not achieved effortlessly. Take the way we acquire a tool or a probe. If, as seeing them, we are blindfolded, we cannot find our way with a stick as skillfully as a blind man

> does who has practiced for a long time. We can feel that the stick hits something from time to time but cannot correlate these events. We can learn to do this only by an intelligent effort at constructing a coherent perception of the things hit by the stick. We then gradually cease to feel a series of jerks in our fingers as such – as we still do in our clumsy trials – but experience them as the presence of obstacles of certain hardness and shape, placed at a certain distance, at the point of our stick. We may say, more generally, that by the effort by which I concentrate on my chosen plane of operation I succeed in absorbing all the elements of the situation of which I might otherwise be aware in themselves, so that I become aware of them now in terms of the operational results achieved through their use. When the new interpretation of the shocks in our fingers is achieved in terms of the objects touched by the stick, we may be said to carry out unconsciously the process of interpreting the shocks. And again, in practical terms, as we learn to handle a hammer, a tennis racket or a motor car in terms of the situation we are striving to master, we become unconscious of the actions by which we achieve this result.[5]

One would have to first construct a coherent perception of things by choosing an illuminating term that helps one to interpret the experience. The purpose of constructing a coherent perception of things has much to do with the motivation of the agent to gain understanding, by using a term that he or she thinks best fits the scenario being described. Once the term has been used and becomes a part of the common idiom (in our case, the academic idiom) the use of this becomes seemingly unconscious although the motivation behind using the term was first conscious.

If one were to find merit in this chain of reasoning, the centrality of anthropomorphizing to the claim *that* the discourse of mind is also a rhetorically constructed one - can be argued. Symbolic AI starts with the assumption that human thinking is based on a language of thought, which employs the operations of mathematical reasoning. By looking at the rhetorical situation that facilitates the emergence of such an assertion, it becomes apparent that the semblance between algebra and symbol-manipulation of the mind was noticed and developed. Algebraic or Boolean logic in turn played an instrumental role in shaping the digital computer. So what is the anthropomorphizing in symbolic AI? The answer is not simple. First, it implies that the language of mathematical logic is the language of human thinking, and this language is most suited for making machines perform tasks that require intelligence. Second, it implies that the digital computer can be seen as a cognitive metaphor for mind. Therefore, the discourse of symbolic AI is anthropomorphic in that mathematical reasoning is considered the language of thought, and that by creating a machine that employs mathematical reasoning; we are essentially creating models of ourselves. Therefore the machine metaphor can be used as a substitute for human cognition if one can consent to the claim that human mental processes are also mechanistic. Hence, the anthropomorphic selection of the machine metaphor stems from the motivation to see the world through the lens of mathematical reasoning. The discourse of connectionist AI replaces the machine metaphor with the brain metaphor by asserting that the biological brain provides a more accurate picture of cognition than the serial computer. Instead of mechanizing or digitizing the brain (as seen in the computational psychology of symbolic AI), connectionist AI seeks to 'biologize' the machine by using the parallel-processing brain as an analog for the machine. The anthropomorphizing is more explicit in connectionism in that the biological brain (and we are biological creatures) is the prototype for artificial nets. The modeling of artificial nets in accomplishing certain tasks are then used as descriptive indicators of mental processing taking place in the brain with regard to the accomplishment of the task at hand. What is perhaps ironic is not so much the replacement of the machine metaphor with a brain metaphor, but the replacement of one type of a machine metaphor with another since the brain is thought by connectionists as a neurological machine. The borrowing of neuroscientific language to describe artificial nets is rhetorical, whereas the motivation to do so is clearly anthropomorphic. These arguments will be addressed in greater detail in chapter 3.

Chapter 4 provides a rhetorical analysis of the work and legacy of Alan Turing. Turing is one of the most influential figures in the history of AI, having significantly influenced both symbolic AI and connectionism. This chapter will trace the rhetorical situation that facilitated the development of Turing's thought and analyze the terministic screens inherent in Turing's idea of thought. While Chapters 2 and 3 primarily serve to introduce the theoretical background for the analysis, the discussion on Turing will serve to advance the first part of the argument that AI is rhetorically constructed by showing how Turing's conception of thought is both historically and rhetorically constructed. Turing's thought is historical insofar as the rhetorical situation is concerned and rhetorical when viewed from the perspective of a terministic screen.

Chapter 5 traces the rhetorical situation behind the development of symbolic AI and the emerging terministic screens from the discourse of mind. The rhetoric plays out in the form of a metaphoric argument 'the mind is a computer' that gives rise to further screens and concepts. From a rhetorical perspective, a terministic screen is a trained incapacity in that our way of seeing the world also becomes a way of not seeing the world. By wearing the lens, namely that the mind is a lens, the question to ask from a rhetorical perspective pertains to the insights and blindspots of computational psychology. The role that the machine metaphor plays in generating a theoretical matrix of computational psychological terms that are in turn used to describe how the mind works. The metaphoric screen rhetorically constructs the conceptual vocabulary in describing mind being significant in both its emphases and omissions. Chapter 6 explores both the rhetorical situation and the terministic screens that come into play when the metaphor shifts from mechanistic to biological. The use of neuroscientific language and the concomitant emphasis on neural realism in the modeling of artificial neural nets facilitates the rhetorical construction of connectionist models of cognition. Paralleling the discussion on symbolic AI in chapter 5, this chapter will look at both the insights and blindspots from a connectionist terministic screen. Both chapters 5 and 6 explore in greater detail how the discourse of mind in both symbolic and connectionist AI are rhetorically constructed, by expounding on the role of generative metaphors and practical reasoning.

Intended Audiences

The overall goal of this discussion is to provoke a conversation with multiple audiences. Parsing the expression "multiple audiences" is perhaps necessary. On one hand, there is a need for reengaging the discipline in an old yet painfully question pertaining to the relationship between words and things. In order to break out of the familiar framing of this issue, another way of attacking this question is to pay attention to a synonymous question – 'what is the nature of thought?' Related to this, is the question 'how are models of thought constructed?'

Second, I would like to engage the AI communities by urging them to self-reflexively examine their own tacit mental constructions, acknowledging both the insights and blind spots. Without dampening the epistemological optimism of scientific realists, a constructivist approach can be beneficial in that, it teases out the equivalent roles that linguistic description and human agents play in the legitimization of human knowledge. Conceiving rhetoric as epistemic is broad enough to encompass persuasion as well, even while acknowledging how language can shape understanding as well. The nexus between rhetoric and epistemology can be evidenced from the participation of linguistic agents shaping both the direction of the discourse and the terms of understanding.[6] It would also be instructive to engage rhetoricians and sociologists of science who are interested in making the connection between metaphor and theory-generation, and to explore how metaphors generate theories.

Symbolic AI

John Haugeland uses the term "good old-fashioned AI," or GOFAI to refer to the classical or "symbol manipulation" or even "language-of-thought" AI. Symbolic AI was prevalent in the mid-fifties through at least the mid-eighties. In Haugeland's words, Symbolic AI suggests that "the mind just *is* a computer with certain special characteristics – namely, one with internal states and processes that can be regarded as explicit *thinking* or *reasoning*."[7] The notion of a formal system is widespread in mathematics, chiefly derived from arithmetic and algebraic systems. By manipulating tokens according to definite rules, people solve arithmetic or algebraic problems. Symbolic AI is "predicated on the idea that systems can be built to solve problems by reasoning or thinking them through in this way, and, moreover, that this is how people solve problems."[8]

Newell and Simon theorized that the human brain and the digital computer,although totally different in structure and description, have a common functional description at a certain level of abstraction. At this abstract level both the human brain and the appropriately programmed digital computer could be seen as two different instantiations of a single species of device, namely a device that is capable of generating intelligent behavior by manipulating symbols by means of formal rules. Newell and Simon thus formulated the Physical Symbol hypothesis as follows:

> A physical symbol system has the necessary and sufficient means for general intelligent action. By necessary we mean that any system that exhibits general intelligence will prove upon analysis to be a physical symbol system. By sufficient we mean that any physical symbol system of sufficient size can be organized further to exhibit general intelligence.[9]

Human learning is primarily seen as symbol-manipulation, following a set of discrete rules. Therefore, intuition, insight and learning are thus no longer exclusively human processes, but any large high-speed computer can be adequately programmed to exhibit them. Newell and Simon believed that the range of problems that the computer could handle would soon become coextensive with the range of problem solving that the human mind is used to. The key here lies in the shift from a purely algorithmic approach to a more heuristic one – heuristic rules are those that rely on plausible solutions to problems, and thus rely on experience of judgment.

At this juncture it is noteworthy to observe both Searle and Dreyfus' critiques of strong AI. AI research has traditionally been bifurcated in terms of two streams, weak AI and b) strong AI. According to weak AI, the computer is a valuable tool in the study of the mind. According to strong AI, the programmed computer is capable of displaying cognitive states and has a mind. Strong AI claims that computers can simulate human ability, even in tasks such as understanding stories. Strong AI claims that the machine can literally be said to *understand* the story and provide answers to questions; and what the machines and its program do *explains* the human ability to understand the story and answer questions about it. John R. Searle impugns the adequacy of the criteria employed by Turing Test, by counter-asserting that computers do not possess the same level of cognition or understanding that its human counterparts seem to display. In response to Turing's imitation game, Searle comes up with his "Chinese Room" thought-experiment whereby he conjures up a situation in which he takes on the role of a formal symbol manipulator, and thus deceive the interrogator into believing that he knew Chinese. Searle explains, "as far as the Chinese is concerned, I simply behave like a computer; I perform computational operations on formally specified elements. For the purpose of the Chinese, I am simply an instantiation of the computer program."[10] Searle extends this analogy to computers, implying that computers may not understand a lick of Chinese, but could still achieve a verisimilitude of cognizance by formally manipulating the symbols, thereby conveying the impression that it knows what it is doing. Thus, in turn, Searle proposes the following:

> If strong AI is to be a branch of psychology, it must be able to distinguish systems that are genuinely mental from those which are not. It must be able to distinguish the principles on which the mind works; otherwise it will offer us no explanations of what is specifically mental about the mental. And the

> mental/nonmental distinction cannot be just in the eye of the beholder – it must be intrinsic to the systems. For otherwise it would be up to any beholder to treat people as nonmental and, for instance, hurricanes as mental, if he likes.[11]

Another vocal critique worth mentioning is that of the Dreyfus' brothers. Dreyfus' critique of strong AI revolves around the non-axiomatic, non-formalizable nature of common sense knowledge.[12] Humans solve problems not so much with the help of technical rationality, but by exercising their intuition and common sense. Technical rationality is a characteristic of formal, mathematical systems. Dreyfus considers it a reductionist move to use the "machine" metaphor as a model to study the mind and vice-versa.

Last but not the least, another important piece worth looking at is Daniel Dennett's notion of an intentional strategy whereby machines are ascribed with beliefs and desires not unlike rational actors.[13]

Connectionism

Connectionism, in layperson's terms, is an approach in artificial intelligence and cognitive science aimed at producing biologically realistic models of the brain and mental processing. The technical term used for this approach is PDP (Parallel Distributive Processing). PDP necessitates a paradigmatic shift from viewing the mind as a symbol manipulator to one in which the brain is considered as a complex intermixture of multilayered networks. Not unlike the neurons in the physico-chemical brain, the units of a network are construed as simple processors, and the seemingly infinite connections between these processors serve as tools to carry on the task of information-processing. In Haugeland's words, "connectionist networks are inspired to some extent by brain and neural networks. The active units are like individual neurons, and the connections among them are like axons and dendrites along which electro-chemical "pulses" are sent from neuron to neuron."[14] The information-processing that takes place here, in the absence of a central processor, is parallel and distributed. The underlying impulse is to look at the brain as a parallel computational device instead of a traditional serial computer.

Paul M. Churchland in his essay "On the Nature of Theories: A Neurocomputational Perspective" begins his piece by discounting the classical view of theories, what he in other words describes as "sentential epistemologies," reminiscent of the GOFAI tradition.[15] Sentential epistemologies derived from logical propositions have confined themselves to aspects such as prediction, explanation, deduction and so forth. Most sentential epistemologies were based on a rationalistic account of the world, which relied on sentences and propositions. In Churchland's words,

> if theories are just sentences, then the ultimate virtue of a theory is truth. And it was widely expected that an adequate account of rational methodology would reveal why humans must tend, in the long run, towards theories that are true....I have been motivated primarily by the pattern of the failures displayed by that approach. Those failures suggest to me that what is defective in the classical approach is its fundamental assumption that language-like structures of some kind constitute the basic or most important form of representation in cognitive structures, and the correlative assumption that cognition consists in the manipulation of those representations by means of structure-sensitive rules.[16]

Thus, in place of a sentential epistemology Churchland suggests an alternative route, namely a neurocomputational perspective on cognition. A neurocomputational perspective can be seen as a rhetorical transaction whereby there is a direct borrowing of terms from neurosciences, in such a manner that the distinction between the organic brain and the artificial net can be conflated and sustained within this model.

With neural realism as its manifest goal, connectionists begin on the grounds of linguistic exactness by fixing the language of neurosciences as a referent for artificial nets; and the latter could be seen additively as representations of human cognition. Interestingly enough, the quest for neural realism is metaphoric in that language constitutes the primary grounds that enable the crossover to take place. Once the nomenclature has been taken care of, the conflation of models with mental phenomena is considered plausible in spite of existing dissimilarities in processes. The principle of additives or the putting together of a network serves as the grounds for machine cognition; and if this is the case, can human cognition be much different since after all we are neurological machines as well.

The rhetorical dimensions of connectionism, or for that matter many other scientific enterprises, by no means undermine the scientific credibility of a project but only goes to highlight the eisegetical interpolations of linguistic lenses in spite of protestations to the contrary.

Terministic Screens and Burkean Tropes

A terministic screen simply implies "even if any given terminology is a *reflection* of reality, by its very nature as a terminology it must be a *selection* of reality; and to this extent it must function also as a *deflection* of reality."[17] Implicit in the screen are the terminal boundaries delimiting the apprehension of anything beyond the lens, and by the same token, facilitating a sharper vision within the parameters of the lens. Terministic screens serve as converging points where interrelationships are established between a set of selected terms. A terministic screen is a tropological concept, a metaphoric lens primarily linguistic in nature allowing the selection and flourishing of a particular vocabulary accompanied by preempting competing lenses from gaining a foothold in the conversation. If the screen posits the mind as a digital computer, there is a corresponding importation of a computational vocabulary in describing mental processes. If the screen posits the artificial net as a brain-like network, there is a corresponding importation of a neuroscientific vocabulary in describing machine processes.

As Kenneth Burke explains, "not only does the nature of our terms affect the nature of our observations, in the sense that the terms direct the attention to one field rather than to another. Also, many of the "observations" are but implications of the particular terminology in terms of which the observations are made. In brief, much that we take as observations about "reality" may be but the spinning out of the possibilities implicit in our particular choice of terms."[18] Burke articulates that a terministic screen in terms of "symbolic action is exercised about the necessarily suasive nature of even the most unemotional scientific nomenclatures."[19]

Burke articulates the centrality of terministic screens in human discourse through the following words: "we must use terministic screens, since we can't say anything without use of terms; whatever terms we use, they necessarily constitute a corresponding kind of screen; and any such screen necessarily directs the attention to one field rather than the other."[20] A terministic screen besides providing the grounds for elaborating a conceptual framework, can also serve as persuasive grounds for gaining the intellectual sympathy of larger audiences who although initially skeptical may find resonance with the ideas because of their favorable inclination towards certain notions and terminology elicited by the screen. For instance, the human-machine isomorphism may evoke a great deal of skepticism even among some scientific materialists who may initially hold the view that biology is a precondition for cognition. Under these circumstances, pro-AI enthusiasts may have to find or locate areas of convergence or common ground by emphasizing the materiality of human consciousness and thinking by focusing on principles such as emergent properties and material architecture among others.

A terministic screen grounded in a dramatistic view of language articulates how linguistic schemas can induce frameworks, by the liberal mapping of one set of terms onto another. The schema of a 'machine' metaphor generates the entire field of computational psychology whereas the schema of a 'brain' generates the mapping of neuroscientific language onto artificial nets. How this is the case will be elaborated in greater detail in the subsequent chapters.

In essence, AI is still a relatively new idea that its gradual acceptance by those outside the perimeter can be painstakingly slow. Michael Polanyi argues that for a new scientific idea to gain adherence, "demonstration must be supplemented…by forms of persuasion which can induce a conversion," and "the refusal to enter on the opponent's way of arguing must be justified by making it appear altogether unreasonable."[21] Proponents of a new system "can convince their audience only by first winning their intellectual sympathy for a doctrine they have not yet grasped. Those who listen sympathetically will discover for themselves what they would otherwise never have understood."[22] In a strikingly salient passage that speaks about the rhetorical dimensions of science, Polanyi states:

> Such an acceptance is a heuristic process, a self-modifying act, and to this extent a conversion. It produces disciples forming a school, the members of which are separated for the time being by a logical gap from those outside it. They think differently, speak a different language, live in a different world, and at least one of the two schools is excluded to this extent for the time being (whether rightly of wrongly) from the community of science.[23]

Polanyi's notion of a different language and Burke's terministic screen share a common assumption that discursive communities with their specialized languages and typologies employ the discourse simultaneously as gateways and gatekeepers of the discourse at hand.

The mediation of terministic screens specific to the respective schools of thought can further be illuminated in the light of Burke's master tropes namely – metaphor, metonymy, synecdoche and irony.[24] Metaphors induce perspectives, metonymies can be seen as reductions where a part is taken to represent the whole or the sum of the parts, synecdoches can be seen as representations and irony can be seen as a substitution for dialectic (used in a restricted sense, referring to the interaction between terms).[25]

These tropes are interrelated and are by no means mutually exclusive. In symbolic AI, the 'machine' or 'computer' metaphor serves as a metonymy by reducing all thought to computation, and computation in turn serves as a synecdoche for human cognition, with irony playing out in the interaction between the vocabularies of human and machine cognition respectively. In connectionist AI, the 'brain' metaphor is carried over into the vocabulary of artificial neural nets, with the neural net serving as metonymic reduction of thought primarily in the language of input-output information processing technologies, and the input-hidden-output units model of neural activity serving as a synecdoche for both human and machine cognition, with irony playing out in the appropriation of a neuroscientific vocabulary in the description of parallel-processing networks.

Burke's voiced punctuation on the centrality of metaphor in the development of language and thought is perhaps apt for the discourse of mind and is consistent with the tone of the ensuing discussion although cognitive science punctiliously effaces reference to metaphors:

> Language develops by metaphorical extension, in borrowing words from the realm of the corporeal, visible, tangible and applying them by analogy to the realm of the incorporeal, invisible, intangible; then in course of time, the original corporeal reference is forgotten, and only the incorporeal, metaphorical extension survives.[26]

Locating the metaphors, and how the metaphors extend the conceptual vocabularies of the discourse is important not only to understand the rhetorical dimensions of AI, but also how the rhetoric is produced. Therefore aspects of textual production are as important as matters of textual reception; and it is precisely this motivation that makes rhetorical perspectives heuristic in understanding not only the presumptions of the discourse but also the presuppositions that enables the production of the discourse.

Undertaking a rhetorical analysis of AI is useful, since a close reading of symbolic AI and connectionism informs our understanding of the strategies, generative metaphors, types of arguments that researchers have used in viewing the mind. And how these metaphors and arguments have in turned played an influential role in constructing cognitive models of human and machine cognition. Te discourse of Artificial Intelligence has become synonymous with the discourse of mind. Or even more simply, the way we talk about intelligent machines is increasingly becoming reflective of the way in which we talk about ourselves.

Artificial Intelligence in Perspective

John Haugeland perhaps provides a good definition for Artificial Intelligence for our purposes, since the definition sheds sufficient light on AI's philosophy of mind:

> The endeavor to understand *mind* (thinking, intellect) in terms of its *design* (how it is built, how it works). It amounts, therefore, to a kind of cognitive psychology. But it is oriented more toward structure and mechanism than toward correlation of law, more towards the "how" than the "what," than is traditional empirical psychology. An "experiment" in mind design is more often an effort to *build* something and make it work, than to observe or analyze what already exists. Thus, the field of Artificial Intelligence (AI), the attempt to construct intelligent artifacts, systems with minds of their own, lies at the heart of mind design. Of course, natural intelligence, especially human intelligence, remains the final object of investigation, the phenomenon eventually to be understood. What is distinctive is not the goal but rather the means to it. Mind design is *psychology by reverse engineering.*[27]

In essence, constructing intelligent artifacts imply constructing intelligent minds. If minds can be built, they are physical entities as well. Therefore, Artificial Intelligence cuts at the joints of Cartesian metaphysics whose substance dualism privileges mind over matter. From the perspective of AI, mental phenomena such as consciousness should be de-essentialized from metaphysics and reconsidered as emergent properties of neurological processes in the human information processing system. The human mind should be seen as a neurological machine if one were to establish equivalencies between humans and machines. A salient example descriptive of the ontological shift from a humanistic 'essentialism' to a more clinical reinterpretation of human behavior in purely mechanistic terms is well-illustrated in the following passage:

> From the behaviorist point of view, what are the fundamental differences between animals and robots? Apart from the obvious, though behaviorally trivial differences – such that they are made from different hardware – we can say that animals are more sophisticated, have better sensors etc. As scientists, we are not inclined to say that robots are machines whereas animals are not, or that the behavior of robots is predetermined whereas that of animals is not. As scientists, we believe that a deterministic philosophy applies to both animals and robots, and this attitude enables us to entertain the notion robots, may, one day, be analogs of animals.[28]

For some AI researchers, the analogy between biological systems and machines is rhetorically established by asserting an equivalence between "machine instructions" and "amino acids" because they "actively manipulate bits, bytes, CPU registers...they are considered analogous to the RNA world, because the same structures bear the 'genetic' information and carry out the 'metabolic' activity."[29]

In effect, the convergence of revolutionary developments in biology, physics and computer science makes Artificial Intelligence a hot-button topic. Concomitantly, the trend has been increasingly moving towards a deterministic view of human behavior. However, such a view is not necessarily uniform across the sciences since it appears that quantum physicists at least, in principle, is willing to embrace partial indeterminacy since the fundamental subatomic particles that constitute the physical world can take on both wave and particle forms. On the other hand, the resurgence of biological discoveries such as the mapping of the human genome have coincided with tremendous breakthroughs in computation rhetorically signifying a return to a form of determinism, far more sophisticated than clockwork mechanics. In some scientific circles, the pervasive culture of scientific optimism is based on the assumption that there can be no ignorabimus.

Rhetoric of Artificial Intelligence

British mathematician and one of the pioneers of Artificial Intelligence, Alan Turing outlined a rhetorical manifesto for machine intelligence.[30] The imitation game, otherwise known as the Turing's test, was put forward as a sufficient index to gauge machine intelligence. The imitation game is played with three people, a human (A), a computer (B) and an interrogator (C) who is obviously human. The interrogator does not know which terminal the machine or the human is operating from. During the course of the conversation, if the machine's responses are indistinguishable from that of the human's, the computer is said to have passed the Turing test and is considered intelligent. The imitation game is based on the idea of mimicking human behavior, to demonstrate that intelligence can be replicated in non-human systems as well. Turing's suggestion is noteworthy "if one wants to make a machine mimic the behavior of the human computer in some complex operation one has to ask him how it is done, and then translate the answer into the form of an instruction table. Constructing instruction table is usually described as programming."[31] A superficial reading of the sense in which Turing employs the term "imitation" may yield the notion that human intelligence is the *ultimate* model. Such a reading will understate the claims of computational psychology, whereby information processing in the human brain is seen as algorithmic and following a set of finite rules. Therefore, computational psychology posits a dialectical interrelationship between humans and machines since the two at a functional level of isomorphism, cognitively share an instrumental view of intelligence enacted by a 'program.'

Andrew Feenberg articulates the view that AI "inspired a new field in psychology which takes the computer as the model of the mind," consistent with the predominantly rationalist outlook of (post)-industrialized societies in which digital operations are seen as the spitting image of thought processes.[32] The manner in which we talk about AI has become isomorphic with self-reflexive talk about human nature. The discourse about AI not only propels the science behind AI, but also rhetorically constructs the ambiance of social acceptability that facilitates a paradigm shift from an essentialized, metaphysical conception of human nature to a digitized post-humanity. The manner in which the nature of human cognition is linguistically framed and the rhetorical nature of the typologies employed by research programs provide common *topoi* that scholars of communication can explore.

Kenneth Burke conceives rhetoric as "an essential function of language itself, a function that is wholly realistic, and is born anew; the use of language as a symbolic means of inducing cooperation in beings that by nature respond to symbols."[33] Perelman suggests that "all argumentation aims at gaining the adherence of minds, and, by the very fact, assumes the existence of an intellectual contact."[34] A rhetorician must seek to engage in argument, a person who "must attach some importance to gaining the adherence of his interlocutor, to securing his assent, his mental cooperation. It is accordingly, sometimes a valued honor to be a person with whom another will be in discussion."[35]

The field of argumentation is dynamic and constantly evolving, covering new epistemological grounds in AI circles. Believers in AI must secure the assent of skeptics and reinforce the convictions of believers on one hand; and also compete with alternative paradigms and visions of how their goals are to be accomplished. AI research is propelled by countervailing conceptions of mind, with the symbolists arguing that mental operations are primarily algorithmic and based on symbol-manipulation while the connectionists argue that artificial neural nets are more realistic than symbol-manipulation. The types of arguments, analogies, metaphors and linguistic strategies are amenable to rhetorical scrutiny. Indeed, the very deployment of suasory terms such as "Artificial Intelligence" and "Mind Design" are indicative of a progressive movement, a definite *telos* towards which these programs are orientated. The strategic employment of words in defining these programs is essentially self-fulfilling, in that both these terms are positive, not so much along metaphysical or essentialist lines, but with regard to what they socially signify. John Poulakos's postulation of rhetoric as the "art which seeks to capture in opportune moments that which is appropriate and attempts to suggest that which is possible" sheds light on the role that the strategic use of language plays in constructing reality.[36]

The appeal to the human mind as a prototype is an isomorphic/anthropomorhic strategy to rhetorically construct computational models that are in turn valorized as realistic depictions of human minds. The symbolic tradition considers thought to be a result of symbol-manipulation, capitalizing on the twin ideas of formal systems and technical rationality. Connectionism sees such a move as an *a priori* top-down imposition of a theoretical framework that is not grounded in neural realism; instead it argues that recapturing neural processes of the brain through biologically realistic models yields better knowledge of cognition. Both symbolic and connectionist AI represent diverging conceptions of the mind, and these rival conceptions have placed these research programs on entirely different paths, although the goal seems to be the same. Underlying both symbolic and connectionist AI is the inevitable sense that researchers use both the mind/brain and accompanying computational systems as models that are reflexive of each other. Although, both systems are entirely different, chiefly at the physico-chemical level there is an isomorphic basis on which the mapping is said to occur. Attempts to portray an accurate picture of how the mind and the machine work mutually shed light on each other. If we subscribe to a model of scientific realism, namely that discourse is a translucent window on the object of its inquiry, then it is imperative to take the claims of either symbolic or connectionist AI at face-value. Such a position can be fitted into Jurgen Habermas' view that critical discourse and technical discourse do not mix since technological discourse is for the most part neutral, except for the fact that it constitutes purposive-rational action.[37] The formal systems foregrounded by technical discourse are built on the pillars of technical rationality, objectivity and consistency. As Feenberg explains, in formal systems "objects are conceptualized as fixed and frozen, unchanging in themselves but available for manipulation from above."[38] Directly antithetical to Habermas' view is the antecedent position promulgated by Herbert Marcuse who states:

> The hypothetical systems of form and functions becomes dependent on another system – a pre-established universe of ends, in which and *for* which it develops. What appeared extraneous, foreign to the theoretical project, shows forth part of its very structure (methods and concepts); pure objectivity reveals itself as object *for a subjectivity* which provides the Telos, the ends. In the construction of the technological reality, there is no such thing as a purely rational scientific order; the process of technological rationality is a political process.[39]

Marcuse's characterization of rationality as a political process might seem misguided and even wrong-headed given the fact that most scientists are by themselves not actively consorting with any political order or even hegemonic capitalism; although capitalism does play a huge role in the furtherance of various scientific disciplines. And none of the literature pertaining to mind design surveyed in this discussion hints any sort of association with any sort of hidden political ideology whatsoever. Even so, one way to make sense of Marcuse's scathing indictment of technology is in the (un)witting transposition of means as ends prevalent in technical discourse.

Either way, whether one were to invoke the high grounds of scientific realism or the radical constructivist position of Marcuse what is germane to a rhetorical perspective, is the mode of representation. Representations are best encapsulated in scientific models; which according to the English physicist Norman Campbell are necessary approximations of physical reality.[40] Models are fiduciary agents of scientific epistemologies, being both gateways and gatekeepers in selecting and deflecting the terms of empirical engagement. A rhetoric and philosophy of technology seeks to decipher and contextualize the terms of engagement, and how these terms are instrumental in shaping the ensuing discourse.

The rhetorical dimensions of AI are manifest in the specific intraparadigmatic communicative practices engendered by discursive communities. That the mind is either a problem-space or a neural network is shared by members of the particular paradigm. Such postulations are based on eliminating conceptions of mind, and therefore reinforcing the particular description at hand. Specialized taxonomies are generated within the paradigm, and the paradigm in turn generates new ways of thinking within the framework that is adopted. As Thomas Kuhn explains:

> What characterizes revolutions, is thus, change in several of the taxonomic categories prerequisite to scientific descriptions and generalizations. That change, furthermore, is an adjustment not only of criteria relevant to categorization, but also in the way in which given objects and given situations are distributed among pre-existing categories. Since such redistribution always involves more than one category and since those categories are interdefined, this sort of alteration is necessarily holistic.[41]

Even so, Kuhn's nominalism is by no means radical since he abides by some sort of referentialism that implies the possibility of demonstrating technical results within the community:

> Proponents of different theories are, I have claimed, native speakers of different languages...I simply assert the existence of significant limits to what the proponents of different theories can communicate to each other...Nevertheless, despite the incompleteness of their communication, proponents of different theories can exhibit to each other, not always easily, the concrete technical results available by those who practice within each theory.[42]

Embedded in Kuhn's notions are the unifying role that certain discursive practices play in bringing together adherents of a particular theory or paradigm. Discursive practices besides their obvious grounding in empirical and inductive inferences also create a stock set of linguistic terms that are putative descriptions of mental realities. Such a gesture is akin to the notion of a 'language game.' Expanding on the notion of a system of differences from phonemics, morphemics and syntax to the actual use of language in the social world, Wittgenstein propounds the idea of a language game. A language game is based on the rules that we employ to govern the use of our language:

> But how many kinds of sentences are there? Say assertion, question and command? - There are countless kinds: countless different kinds of use of what we call "symbols," "words," "sentences." And this multiplicity is not something fixed, given once for all; but new types of language, new language-games, as we may say come into existence, and others become obsolete and forgotten...Here the term "*language-game*" is meant to bring into prominence the fact that the *speaking* of language is part of an activity, or a form of life.[43]

The crux of Wittgenstein's argument is along the following lines, "if language is to be a means of communication there must be argument not only in definition but also in judgments."[44]

The language game in Symbolic and connectionist AI are metaphorically mediated. These metaphors are conceptual in nature in that they consist of a "conceptual mapping of entities, properties, relations and structures from a domain of one kind (the source domain) onto a domain of a different kind (the target domain).[45]

Diego Fernandez-Duque and Mark L. Johnson state that the mind as an "information processing device metaphor" in AI yields the following:

Source Domain (Communication System)	*Target Domain (Mind)*
Transmitter	Information Source
Input Modules (Buffers)	Iconic Memory
Parallel Processing Channels	Sensory Systems
Filter	Attention
Input	Stimulus
Signal	Target Signal
Noise	Distractors.[46]

Symbolic AI's viewing the mind as a computer helps one to borrow machine vocabulary in order to create a computational psychology of mind (Chapter 5). The mind as machine is an attention or orientation metaphor that generates conceptual terms to describe human psychology in terms of computational psychology. If the hardware of the 'mind' is comparable to a computer, then human thinking is similar to the instantiation of a computer program. The computer program is considered representative of human thinking and from an isomorphic perspective, symbol manipulations are seen as representations of thinking processes. If the mind is an information processing system, information is designated by certain symbols that follow certain rules. Symbols are the designations of distinct objects, and an information processing system is said to contain "a collection of these symbol structures" that serve as data structures.[47] Besides data structures, the system is also said to have a collection of processes that operate on expressions to produce other expressions: "processes of creation, modification, reproduction, and destruction."[48] Philip E. Agre articulates that from the use of computational vocabulary two patterns emerge:

> A word that once referred to something in the world now refers to a structure in the computer. Common examples include "situation," "pattern," "context," "object," "list," "map," "structure," and "problem." Individual AI researchers have defined hundreds of others.
> A word that once referred to an activity conducted by agents in the world now refers to a process occurring entirely in the computer. Examples include "search," all verbs for operations on data structures ("construct," "manipulate," "inspect," "point at," "transverse," "collect," "recycle") and many predicates on the internal operations of technical entities.[49]

If the computer analogy is to be transported into human information processing, the following would be most likely to happen: "if agents need to think about the world, put analogs of the world in the head. If agents need to act in situations, put data "structures" called situations in the head. If agents need to figure out what might happen, put simulations of the world in the head. The tacit policy…is to reproduce the entire world inside the head."[50]

In the computational realm, all such activity is said to take place inside the computer. Therefore when cross-domain mapping is said to occur from the source to the target domains, the linguistic and conceptual vocabulary of the source domain is transported to the target domain, such that the target domain (human mind) is redescribed in terms of the source domain (the machine or computer). Inside the source domain, symbolic processing is said to occur, and thinking requires the manipulations of symbols with a very specific structure, the structure of a formalized language. Computational psychology operates under the assumption that the mind "uses a formalized language (or something like a formalized language) both as medium of computation and medium of representation."[51] Symbol manipulation via a formalized language is said to represent thinking. As Hilary Putnam explains:

> I believe we cannot account at all for the functioning of thought and language without regard to at least some mental items as representations. When I think 'there is a tree in front of me,' the occurrence of the word 'tree' in the sentence I speak in my mind is a *meaningful* occurrence and one of the items in the *extension* of that occurrence of the word 'tree' is the very tree in front of me. Moreover, the open sentence '*x* is in front of me' is correlated (in the correct semantics for my language) with the relational property of being in front of me, and the entire sentence 'there is a tree in front of me' is, by virtue of these and similar facts, one which is *true* if and only if there is a tree in front of me.[52]

Therefore, in computational terms, the tree is represented by a string of symbols, which constitutes the machine's language and acted upon by a high-level language which constitutes the set of instructions. The mode of representation in computers are said to be analogous to representations in the human mind rendering credibility to human-machine isomorphisms. In the source domain, 'thinking' is said to take place through a series of automated subroutines, therefore in the target domain, mental operations "involves series of sequential automated subroutines."[53]

From a rhetorical perspective, what is salient is the reinterpretation of mental operations in terms of a distinct computational vocabulary. To test this, one only has to consult textbooks in psychology before the popularization of the computer metaphor. The adventitious use of modern technological artifacts provides a rich repository of metaphorical armaments for philosophers and psychologists who are in the business of (re)interpreting human nature.

With connectionism, we see a distinct shift in that what is the source domain in symbolic AI becomes the target domain and vice-versa. The human brain becomes the source of inspiration for building artificial nets. The importation of neuroscientific language into the realm of artificial nets, recasts machines in biological terms. Therefore artificial nets are seen as the formal equivalents of the parallel processing brain. Computers are like brains. Interestingly enough, the shift in metaphor from a mechanistic to a biologistic one is accomplished purportedly for the sake of achieving neural realism (Chapter 6). Neurons are seen as the fundamental processing unit or atoms of verbal behavior. The application of the 'brain' metaphor takes the following dimensions:

Source domain (Brain)	Target Domain (Machines)
Neuron	artificial neurons
Neural synapses	artificial connections
Human vision	pattern recognition
Human learning	modification of connection weights
Human communication channel	transmitter-channel-modules
Nervous system	connectionist networks

Neuroscientific language is liberally employed to describe connectionist networks, such that the neural modeling of artificial nets are then turned around to describe human cognition. However, the shift from the mechanistic to the biologistic metaphor does not preclude mechanistic dimensions in neural networks. For instance, the language used in communication technologies has been freely borrowed in describing nervous activity. Claude Shannon and Warren Weaver describe communication in terms of a vocabulary borrowed from information transmission technologies: "When I talk to you, my brain is an information processing source, yours the destination; my vocal system is the transmitter, and your ear and the associated eight nerve is the receiver."[54] The nervous system is seen as a "channel for communication," and it "acts to some extent as a single communication channel."[55] The language of broadcast models of communication with its emphasis on Sender-Channel-Receiver and Input-Output language, is used not only to describe human communication but the nervous system in itself. Broadcast models of communication derive much of their inspiration from the invention of radio and television technologies, with their top-down, Sender-Receiver paradigms of communication. Using this language to describe human brains helps legitimize the view that the brain is a neurological machine. Therefore, the shift from the mechanistic metaphor to a biological brain is only a shift from one type of a mechanistic view to another.

Notes

[1] G. K. Chesterton, *The Defendant* (New York: Free Port, 1972), 84.
[2] Daniel E. Gershenson and Daniel A. Greenberg, *Anaxagoras and the Birth of Physics* (New York: Blaisdell, 1964).
[3] Zenon W. Pylyshyn, "Metaphorical imprecision and the "top-down" research strategy," in *Metaphor and Thought, Second Edition*, ed. Andrew Ortony (Cambridge: Cambridge University Press, 1993), 543-558 (553). This is only one example where the basic argument is made that scientific discourse is largely nonconstructivist.
[4] Lloyd F. Bitzer, "The Rhetorical Situation" in *Contemporary Rhetorical Theory: A Reader*, Ed. John Luis Lucaites, Celeste Michelle Condit and Sally Caudill (New York: Guilford Press, 1999), 217-225 (217).
[5] Michael Polanyi, *Personal Knowledge: Towards a Post-Critical Philosophy* (Chicago: The University of Chicago Press, 1974), 61.
[6] Robert L. Scott, "On Viewing Rhetoric as Epistemic," *Central States Speech Journal* 27 (1967): 258-266.
[7] Hagueland, *Mind Design II*, 16.
[8] Hagueland, *Mind Design II*, 19-20.
[9] Allen Newell and Herbert A. Simon, "Computer Science as Empirical Inquiry: Symbols and Search," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: MIT Press, 1997), 81-110 (87).
[10] John R. Searle. "Minds, Brains and Programs," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: MIT Press, 1997), 183-204 (185).
[11] Searle, "Minds, Brains and Programs," 191.
[12] Hubert L. Dreyfus, "From Micro-Worlds to Knowledge Representation: AI at an impasse," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: MIT Press, 1997), 143-182.
[13] Daniel C. Dennett, "True Believers: The Intentional Strategy and Why it Works," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: MIT Press, 1997), 57-80.
[14] Haugeland, "What is Mind Design?," 22.
[15] Paul M. Churchland, "On the Nature of Theories: A Neurocomputational Perspective," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: MIT Press, 1997), 251-292.
[16] Churchland, "On the Nature of Theories," 252-253.
[17] Kenneth Burke, *Language as Symbolic Action: Essays on Life, Literature and Method* (Berkeley: University of California Press, 1973), 45.

[18] Kenneth Burke, *Language as Symbolic Action: Essays on Life, Literature and Method* (Berkeley: University of California Press, 1975), 46.
[19] Burke, *Language as Symbolic Action*, 45.
[20] Burke, *Language as Symbolic Action*, 50.
[21] Michael Polanyi, *Personal Knowledge: Towards a Post-Critical Philosophy* (Chicago: The University of Chicago Press. 1974), 151.
[22] Polanyi, *Personal Knowledge*, 151.
[23] Polanyi, *Personal Knowledge*, 151.
[24] Kenneth Burke, *A Grammar of Motives* (Berkeley: University of California Press, 1969).
[25] Burke, *A Grammar of Motives*, 503.
[26] Burke, *A Grammar of Motives*, 506.
[27] John Haugeland, "What is Mind Design?," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: MIT Press, 1997
[28] David J. McFarland, "Animals as Cost-Based Robots," in *The Philosophy of Artificial Life*, ed. Margaret A. Boden (Oxford: Oxford University Press, 1996), 179-205 (179).
[29] Thomas S. Ray, "An Approach to the Synthesis of Life," in *The Philosophy of Artificial Life*, ed. Margaret A. Boden (Oxford: Oxford University Press, 1996), 111-145 (114).
[30] Alan M. Turing. "Computing Machinery and Intelligence," In *Mind Design II*, ed. John Haugeland. (Cambridge, Massachusetts: MIT Press, 1997), 29-56.
[31] Turing, "Computing Machinery and Intelligence," 34.
[32] Andrew Feenberg, *Critical Theory of Technology* (Oxford: Oxford University Press, 1991), 97.
[33] Kenneth Burke, *A Rhetoric of Motives* (Berkeley: University of California Press, 1969), 43.
[34] CH. Perelman and L. Olbrechts-Tyteca, *The New Rhetoric: A Treatise on Argumentation* (Notre Dame: The University of Notre Dame Press, 1969), 14.
[35] Perelman and Olbrechts-Tyteca, *The New Rhetoric*, 16.
[36] John Poulakos, "Towards a Sophistic Definition of Rhetoric," in *Contemporary Rhetorical Theory*, ed. John Luis Lucaites, Celeste Michelle Condit and Sally Caudill (New York: Guilford Press, 1999), 26.
[37] Jurgen Habermas, "Technology and Science as Ideology," in *Toward a Rational Society*, trans. J. Shapiro (Boston: Beacon Press, 1970).
[38] Feenberg, *Critical Theory of Technology*, 170.
[39] Herbert Marcuse, *One-Dimensional Man* (Boston: Beacon, 1964), 168.
[40] Norman Campbell, *Foundations of Science* (New York: Dover, 1957).
[41] cited in Ian Hacking, *Representing and Intervening* (Cambridge: Cambridge University Press, 1983), 109.
[42] cited in Hacking, *Representing and Intervening*, 13.
[43] Ludwig Wittgenstein, *Philosophical Investigations*, trans. G.E.M. Anscombe (Oxford: Basil Blackwell, 1958), 23.
[44] Wittgenstein, *Philosophical Investigations*, 242.
[45] Diego Fernandez-Duque and Mark L. Johnson, "Attention Metaphors: How Metaphors Guide the Cognitive Psychology of Attention," in *Cognitive Science* 23 (1999): 83-116.
[46] Fernandez-Duque and Johnson, "Attention Metaphors," 86.
[47] Allen Newell and Herbert A. Simon, "Computer Science as Empirical Inquiry: Symbols and Search," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: MIT Press, 1997), 81-110 (86).
[48] Newell and Simon, "Computer Science as Empirical Inquiry," 86.
[49] Philip E. Agre, *Computation and Human Experience* (Cambridge: Cambridge University Press, 1997), 51.
[50] Agre, *Computation and Human Experience*, 51.
[51] Hilary Putnam, "Computational Psychology and Interpretation Theory," in *Artificial Intelligence*, ed. Rainer Born (New York: St. Martin's Press, 1987), 1-17 (3).
[52] Putnam, "Computational Psychology and Interpretation Theory," 3.
[53] Fernandez-Duque and Johnson. "Attention Metaphors," 85.

[54] Claude Shannon and Warren Weaver, *The Mathematical Theory of Communication* (Urbana: The University of Illinois Press, 1949), 17.
[55] D. Broadbent, *Perception and Communication* (New York: Pergamon, 1958), 297.

CHAPTER II

ARTIFICIAL INTELLIGENCE AS RHETORIC OF SCIENCE

> The rhetorical view of science does not deny "the brute facts of nature,:" it merely affirms that these "facts," whatever they are, are not science itself...What range of "brute facts" is worth investigating? How is this range to be investigated? What do the results of these investigations mean? Whatever they are, the "brute facts" themselves mean nothing; only statements have meaning, and of the truth of the statements we must be persuaded. These processes by which problems are chosen and results interpreted, are essentially rhetorical: only through persuasion are importance and meaning established. As rhetoricians, we study the world as meant by science – Alan G. Gross.[1]

Gross articulates an unapologetic apologia for rhetorical studies of science as follows: "...rhetoric is more than window dressing; it concerns the necessary and sufficient conditions for creation of persuasive discourse in any field. Science cannot be excluded by fiat."[2] Gross' defense of rhetoric should be understood in the light of serious criticism that rhetoricians have no business dabbling with the hard sciences, implicit in the notion "good research needs no rhetoric, only clarity."[3] A rhetorical understanding of science articulates the view that there is indeed an active directing of one's intentions in the act of interpretation.

More specifically, one should ask – 'what must a rhetoric of science accomplish?' John Lyne's response is lucid:

> A mature rhetoric of science should tell us something more concrete about how scientists use language and authority to engage audiences and lure them into sharing their view of things. It should tell us something about the strategies and implications of that engagement, including how understanding is distorted as well as how it is enhanced. In short, it should highlight the ways that scientific discourses are relative and compelling for audiences.[4]

Lyne's conceptualization focuses on how rhetoric induces persuasion and facilitates understanding. In order to illuminate the persuasive and epistemic functions it is necessary to: a) trace the rhetorical situation, namely the contexts that facilitate the development of certain theoretical lenses in technical literature and b) discuss particular linguistic strategies such as the use of metaphors that not only persuade but also generates a conceptual vocabulary or terms that becomes part of the framework with which a specific issue is attacked.

The thesis statement of this dissertation revolves around the statement 'AI is a rhetorically constructed inquiry,' part of which pertains to the presence of generative metaphors. The selection of metaphors is legitimized on the grounds of analogical reasoning. It is safe to say that science relies heavily on analogy, and how scientists use analogy is an area that is of great interest to rhetoricians of science.

The field of scientific argumentation in AI thrives on analogy – be it the 'machine' or 'brain' metaphor. Therefore, in that spirit, the purpose of this chapter is two-fold: a) to discuss AI as a rhetoric of science and b) discuss the importance of analogical reasoning as the grounds on which the metaphors that legitimize human-machine isomorphism operate.

In the next section, this paper will explore how analogical reasoning is a key concept in the rhetoric of science and how it plays out in the discourse of mind design.

Analogical Reasoning, Artificial Intelligence and the Rhetoric of Science

Scientific argumentation relies on analogical reasoning implying a rhetorical model of decision-making and problem solving that looks for a similitude between two entities; the similitude is often established by means of a metaphor. Albert R. Jonsen and Stephen Toulmin articulate that analogical reasoning fills in the lacunae where normative reasoning alone won't work:

> The fact that every moral maxim, rule, or other generalization applies to certain actual situations centrally and unambiguously but to others only marginally or ambiguously, makes the latter situations just as problematic in their own way as the situations in which different rules or maxims come into conflict. To put the central point, in a nutshell, once we move far enough away from the simple paradigmatic cases to which the chosen generalizations were tailored, it becomes clear that no rule can be entirely self-interpreting.[5]

Casuistry thus emerges as a necessity, uncovering the following assumptions:

> Similar type cases ("paradigms") serve as final objects of reference in moral arguments, creating initial "presumptions" that carry conclusive weight, absent "exceptional circumstances.
> In particular cases the first task is to decide which paradigms are directly relevant to the issue each raises.
> Substantive difficulties arise, first, if the paradigms fit current cases only ambiguously, so the presumptions they create are open to serious challenge.
> Such difficulties arise, first, if the paradigms fit current cases only ambiguously, so the presumptions they create are open to challenge.
> The social and cultural history of moral practice reveals a progressive clarification of the "exceptions" admitted as rebutting the initial moral presumptions.
> The same social and cultural history shows a progressive elucidation of the recognized type cases themselves.
> Finally, cases may arise in which the factual basis of the paradigm is radically changed.[6]

Analogical reasoning or casuistry is simply case-by-case reasoning, whereby similar cases are clustered together to derive principles of consistency within similar cases. CH. Perelman and L. Olbrechts-Tyteca differentiate the difference between rhetoric and demonstrative reasoning as follows:

> When the demonstration of a proposition is in question, it is sufficient to indicate the processes by means of which the proposition can be obtained as the final expression of a deductive series, which had its first elements provided by

> the constructor of the axiomatic system within which the demonstration is accomplished. Where these elements come from, whether they are impersonal impersonal truths, divine thoughts, results of experiment, or postulates particular to the author, these are questions which the logician considers foreign to his discipline. But when it is a question of arguing, of using discourse to influence the intensity of an audience's adherence to certain theses, it is no longer possible to neglect completely, as irrelevancies, the psychological and social conditions in the absence of which argumentation would be pointless and without result.[7]

Perelman further elaborates:

> Analogies are important in invention and argumentation fundamentally because they facilitate the development and extension of thought. With the phoros as starting point, they make it possible to give them the theme of a structure and to give it a conceptual setting... The shape given to science in the field of electricity is due to the fact that the comparison of electrical and hydraulic phenomena occasioned developments which can explain, complete, and extend the original analogy.[8]

Analogies serve at least two purposes: a) it helps find common places of argumentation and b) it serves as an extension of thought by transporting the vocabulary of one realm into another. When does analogical reasoning serve best and when is it appropriate? In my opinion, it seems that analogical reasoning is appropriate whenever the certainty with which truth-based claims are made in apodictic reasoning seems elusive. In spite of the ascendancy of the hard sciences as veridical models of reality, the type of evidence gathered and gleaned revolves around the realm of plausibility or probability and not on didactic certainty. Richard McKeon's comments on the differences between demonstrability and probability is pertinent to our discussion:

> A demonstration would be the systematic consequences that follow from hypotheses or postulates set down. An experiment is the concrete indication of circumstances in which this would occur. It's the difference between a proof which is universal – a geometric demonstration is always universal – and the construction of something which would be an example of that and is, therefore, a good approach.[9]

Consistent with the distinction between demonstrative and analogical reasoning (that borders more on rhetoric and practical reasoning), Jürgen Habermas differentiates rational and rhetorical reconstruction as follows: rational construction can be categorized as "the cognitive interest in technical control over objectified processes"[10] whereas rhetorical reconstruction is "oriented toward mutual understanding in the conduct of life," "directed towards the transcendental structure of various actual forms of life, within each of which reality is interpreted according to a specific grammar of world-apprehension and of action."[11] However, even the dissociation between rational and rhetorical that Habermas makes is artificial because rational deliberation most often conceals its own rhetorical dimensions due to the fear that 'rhetoric' might undermine the credibility of the enterprise.

Scientific advocacy orients itself towards finding common places by means of analogy and proceeds largely on utilitarian principles. Analogical reasoning serves as grounds for making inferences especially when it is hard to engage a field on its own terms. The inability to make *ex cathedra* pronouncements reasserts the vitality of analogical reasoning in scientific discourse. The analogy if effective is a rhetorical inducement to create consent and works as a means of justifying as well. Charles Arthur Willard articulates that argumentation in general has social grounds (these grounds could be articulated through analogical reasoning) especially when it comes to the aspect of 'justification:' "Justification is something people do that is worthy of study…it is a sociological practice of great importance to explanations of how things muster as knowledge."[12] Analogical reasoning is a form of casuistry. Jonsen and Toulmin explicate the connection as follows:

> First, casuists attempt to classify the event in question, drawing upon paradigms and taxonomies. Frequently, this process involves analogical reasoning, drawing together similarities and differences between well-established paradigms and novel cases as an initial step towards "getting a handle" on such cases. Second, casuists identify which presumptions are relevant to the event...Third, casuists comment on the case's circumstances and how these might affect our overall judgment of the event in question. Fourth, casuists often reflect upon the opinions of prior authorities as these might bear upon our moral assessment of the case. Fifth, bringing together the materials from the first four components, casuists render a verdict.[13]

Although the exact sequential order in which casuistry is enacted may not be the same for analogical reasoning in the sciences, there is certainly a tendency to emphasize the similarities and explain away the dissimilarities as 'anomalies' that will eventually be accounted for. As Perelman explains,

> Although analogy is reasoning that deals with relations existing within the phoros and within the theme, what distinguishes analogy fundamentally from simple mathematical proportion is that in analogy the nature of terms is never a matter of indifference. For the effect of an analogy is to bring the terms *A* and *C* and *B* and *D* closer together, which leads to an interaction and, more specifically, to increasing or decreasing the value of the terms of the theme.[14]

The interaction of terms leads to a similitude which Richard Weaver describes as "correspondences:"

> Those who argue from similitude invoke essential correspondences...Thinkers of the analogical sort use this argument chiefly. If required to characterize the outlook it implies, we should say that it expresses belief in a oneness of the world, which causes all correspondences to have a probative value.[15]

Weaver enunciates the view that "the community of language gives one access to significances at which he cannot arrive otherwise. To find a single word is to find a meaning; to create a word is to find a single term for a meaning partially distributed in other words."[16] The rhetorical construction of intelligence is an example of analogical reasoning where there is a conscious direction of attention in a particular metaphor that helps fashion the discourse.

The Turing test is an exemplar for analogical reasoning, in which the similarities of interaction between the human computer and the machine serve as a basis for stating that machines are capable of engaging in intelligent behavior. The 'physical symbol hypothesis' is also a form of analogical reasoning whereby one physical symbol system is said to be equivalent with another. Allen Newell and Herbert A. Simon suggest that every general intelligence must be realized by a symbolic system.[17] Newell and Simon articulates symbol systems as viable means of representing situations that generate or require intelligent action:

> The symbol-system hypothesis implies that the symbolic behavior of man arises because he has the characteristics of a physical symbol system. Hence, the results of efforts to model human behavior with symbol systems become an important part of the evidence for the hypothesis, and research in artificial intelligence goes on in close collaboration with research in information-processing psychology...[18]

Symbol systems are considered both paradigms and models by means of which to understand human intelligence and co-extensively to produce machine intelligence as well. Simon's borrows his taxonomy of representation from psychology, whereby situations are represented in terms of models:

> We do not have an exhaustive taxonomy of possible representations, but a few basic forms show up prominently in psychological representations. First, situations may be represented in words or in logical or mathematical notations. All of these representations are basically propositional, and are more or less equivalent to a set of propositions in some formal logic. Propositional representation immediately suggests that the processing will resemble logical reasoning or proof.[19]

The physical symbol hypothesis and the notion of finding equivalence of representation among symbol systems is indicative of a simple but elegant syntactic, representational view of the mind vis-à-vis reality.

Representationalism, according to Newton Garver and Seung-Chong Lee implies at least the following:

> One important feature of the ideas represented by signs is that ideas are timeless, in the sense that they are not to be located and identified spatiotemporally. In this respect they differ from the acts of communication, the actual utterances, that occur in the course of our activities as language-users. That sign represent timeless ideas seems, perhaps misleadingly, confirmed by our using dictionaries in which the meaning of words is generally given without reference to particular temporal contexts.[20]

Scientific communication regularly seeks symbolic equivalence between signs and the things they supposedly signify. Although the physical symbol system hypothesis in itself does not demonstrate how representationalism works, it does operate under the assumption that one could produce intelligent behavior through the use of appropriate symbols and commands.

The connectionists (who will be discussed in greater detail in chapter 6) attempt at least the following: a) explain how the brain *represents* the world, and how it performs computations on these representations, b) explain sensorimotor co-ordination vis-à-vis its relationship with computation and c) explain the microphysical organization of the brain and demonstrate how its structure "*implements* the representational and computational activities that the brain as a whole displays."[21] As John O. Greene argues, "from a cognitive perspective behavior is to be explained by reference to the mental operations which produced it...any (communicative) behavior must have arisen purely as a result of the information processing system; there simply are no other inputs to the efferent system."[22] Robert D. McPhee explains that underlying most cognitive theories of mind there are at least three dominant undergirding assumptions:

> First, they (cognitivists) assume that "the cognitive system" has *integrity* as a distinct system with its own structure and processes, which should be studied as processing-input information in a relatively stable ways. Often they assume that cognitive processes are emergent – on a different level of analysis from the neurophysiological....Second, they assume that cognition is a *causal* system at the cognitive level of analysis. They seek lawlike process regularities, involving cognitive constructs like plan or speed of recall, that depend only on the state of cognitive system being focused on, not on a broader interpretive frame or social process. This assumption may be abetted by the computer metaphor that underlies much cognitive theory – the tacit (and often explicit) equation set up between cognitive structures/processing of a computer program....Third, they assume that the cognitive system is *central* in interaction and its explanation – that programmed processes are the most important mediators of the past and shapers of the future.[23]

Connectionism shares some of the qualities that McPhee mentions, and at the bottom, essentially represents a computational model of cognition using parallel distributed processing. David E. Rumelhart et al explain Parallel Distributed Processing (PDP) models as follows:

> These models assume that information processing takes place...through the interactions of a large number of simple processing elements called units, each sending excitatory and inhibitory signals to other units. In some cases, the units stand for possible hypotheses about such things as the letters in a particular display or the syntactic roles of the words in a particular sentence. In these cases, the activations stand roughly for the strengths associated with the different possible hypotheses, and the interconnections among the units stand for the constraints the system knows to exist between the hypotheses. In other cases, the units stand for possible goals and actions, such as the goal of typing a particular letter, or the action of moving the left index finger, and the connections relate goals to subgoals, subgoals to actions, and action to muscle movements. In still other cases, units stand not for particular hypotheses or goals, but for aspects of these things. Thus a hypothesis about the identity of a

> word, for example, is itself distributed in the activations of a large number of units.[24]

In the connectionist model, thought is treated as computation that occurs through the interaction of axons and dendrites. The representation of neural interaction is construed as the representation of thought or mental activity, since there is a neurological basis for cognition and mental activity. Andy Clark justifies the equivalence of thought and computation as follows:

> Why treat thought as computation? The principal reason *(apart from the fact that it seems to work!* – (emphasis mine) is that thinkers are physical devices whose behavior patterns are reason respecting. Thinkers act in ways that are usefully understood as sensitively guided by reasons, ideas and beliefs. Electronic computing devices show us one way in which this strange "dual profile" (of physical substance and reason-respecting behavior) can actually come about.[25]

Is the brain an input-output machine? Or by the same token, can computers be construed as electronic brains? Philip E. Agre comments that the tremendous success of behaviorism and the mathematization of the cognitive sciences gave the field of connectionist Artificial Intelligence and cognitive psychology, a tremendous boost and an air of respectability:

> Cognitivism and behaviorism…shared the mentalist vocabulary of inside and outside, stimulus and response, contents and behavior. At issue was the question of whether scientific sense could be made of the notion of abstract mental structures and processes: thoughts and thinking, memories and remembering, plans and planning, and a boundless repertoire of other nouns and verbs that shuttle easily between the vernacular and scientific vocabularies of psychology. The conception of computation as implemented mathematics provided the license that cognitivism needed. Make it mathematical, or make it seem likely to add finite formalization, and it becomes a valid psychological category. As the project of AI accelerated, the mind became a space for the free exercise of the theoretical imagination.[26]

The metaphysical and intangible term 'mind' is replaced by a more tangible term 'brain,' that is in turn construed as a finite problem-space that subjects itself to algorithm-like formalization while engaging in mental activity. Mental operations are represented by computational programmed procedures and through a dexterous act of linguistic framing based on practical reasoning and sociological forces the mind is equated and eventually replaced by the term 'computer' as a means of representing mental activity. The language used to describe most mental concepts has largely been derived from "English, formal logic, programming languages, or some composite of these; reasoning that proceeds through a detailed simulation of the outside world or through the calculation of obscure numerical functions; processing organized in a centralized or a decentralized manner."[27] These acts of borrowings are largely metaphoric extensions of thought.

Metaphors in Discourse of Mind

The use of metaphors or the construction of a creative linguistic taxonomy does not take away from the scientific nature of cognitive theories or scientific theories in general. As Richard Boyd elucidates, "if the metaphors are apt, and if they are drawn in sufficient detail, the differences in functional (relational) properties of the literal referents…will serve – by analogy – to disambiguate the referents of these terms in their theory-constitutive metaphorical applications."[28] As stated earlier, the naming of things either through relevant metaphors or a specialized taxonomy does not take away the science from a field, but yet, it does shed insights into the rhetorical manner in which knowledge is constituted.

Metaphoric representations of mind have played a significant role in influencing the discourse of mind, even in precursors to AI. British empiricism played an influential role in shaping the course of post-Cartesian philosophical psychology. The translucent, immaterial mind of Cartesian reasoning slowly gave way to a sensory-based picture of mental perception. John Locke rejected the rationalist view of innate ideas and emphasized the importance of experience as the overarching principle of knowledge:

> *Let us then suppose the mind to be, as we say, white paper devoid of all characters, without any ideas.* How comes it to be furnished? Whence comes it by that vast store which the busy and boundless fancy of man has painted on it, with an almost endless variety? Whence has it all the materials of reason and knowledge? To this I answer, in one word, from experience: in that, all our knowledge is founded, and from that it ultimately derives.[29]

Locke's *tabula rasa* is a metaphoric epistemology of human perception, whereby the mind is portrayed as an enclosed space, both in local and global senses, such that impinging impressions on the corporeal mind creates perception:

> Thus the first capacity of humane intellect is, that the mind is fitted to receive impressions made on it either through the senses by outward objects, or by its own operations, when it reflects on them. This is the first step a man makes towards the discovery of anything, and the groundwork whereon to build all those notions which ever he shall have naturally in this world. All those sublime thoughts, which tower above the clouds, and reach as high as heaven itself, take their rise and footing here: in all that great extent wherein the mind wanders, in those remote speculations it may seem to be elevated with, it stirs not one jot beyond those ideas which sense or reflection have offered for its contemplation.[30]

The mind is seen as an enclosure and a receptacle, whereby ideas are received even while serving as a spatial conduit that reinforces distinctions such as inside and outside, back and front among others: "the understanding is not much unlike a closet wholly shut from light, with only some little opening left, to let in external visible resemblances, or ideas of things without."[31] Locke further elaborates "would the pictures coming into such a dark room but stay there, and lie so orderly as to be found upon occasion, it would very much resemble the understanding of a man in reference to all objects of sight and the ideas of them."[32] Locke's theory of mind entails a delineation of features such as inside, hidden, outside that predates input-output models of cognition.

David Hume's theory of mind possessed metaphorical elements as well:

> The mind is a kind of theater, where several perceptions successively make their appearances; pass, re-pass, glide away and mingle in an infinite variety of postures and situations. There is properly no simplicity in it at one time, nor identity in different; whatever natural propension we may have to imagine that simplicity and identity. The comparison of the theater may not mislead us. They are the successive perceptions only, that constitutes the mind.[33]

In spite of Hume's tendency to value empirical realism over metaphysical descriptions, he employs the rich language of a theatrical metaphor, with its notions such as stage, scene, act among others to describe mental life. A metaphorical description emerges as a process of transference, whereby the linguistic resources of a particular concept are considered heuristically salient to explain another concept. Metaphorical descriptions facilitate inductive generalizations by projecting a particular lens or orientation that facilitates the furtherance of explanations of a particular typology.

Locke and Hume were both scientific realists in that they shared a common disdain for rhetoric and poetic descriptions of reality, even so, their combined distrust in itself did not preclude them from resorting to metaphors when a particular metaphor enhances the explanatory value of their inquiries. Given the fact, that even post-Cartesian scientific psychology with its penchant for realism grasps for metaphors as a means of explanation it is helpful to briefly review what some of the current scholars in metaphors are saying about how metaphors influence accounts of cognition.

It must be mentioned that in spite of a lack of a clear consensus pertaining to the nature of metaphor, a common theme that scholars of metaphor share is the need to go beyond the conventional literal/figurative dichotomy and thereby accord metaphors much more than token acknowledgment in human inquiry. Despite thematic differences, most of the scholars cited here take sides with the constructivist position seeing metaphors as much more than substitutions or comparisons of one term with another, and instead as instruments generating conceptual topographies by further engendering a series of subcategories within the particular model of discourse.

George Lakoff and Mark Johnson stress the role of metaphors, in shaping everyday discourse and the direction of conceptual systems:

> Metaphor is for most people a device of poetic imagination and the rhetorical flourish – a matter of extraordinary rather than ordinary language. Moreover, metaphor is typically viewed as characteristic of language alone, a matter of words rather than thought or action...The concepts that govern our thought are not just matters of the intellect. They also govern our everyday functioning, down to the most mundane details. *Our concepts structure what we perceive, how we get around in the world, and how we relate to other people. Our conceptual system thus plays a central role in defining our everyday realities. If we are right in suggesting that our conceptual system is largely metaphorical, then the way we think, what we experience, and what we do everyday is very much a matter of metaphor.*[34]

Lakoff and Johnson argue that there is an inherent systematicity about metaphors that direct our method of seeing, and subsequently not seeing the world:

> The very systematicity that allows us to comprehend one aspect of a concept in terms of another will necessarily hide other aspects of the concept. In allowing us to focus on one aspect of a concept, a metaphorical concept can keep us from focusing on other aspects of the concept that are inconsistent with the metaphor.[35]

Metaphors can also take on the "orientational" mode: "But there is another kind of metaphorical concept, one that does not structure one concept in terms of another but instead organizes a whole system of concepts with respect to one another."[36] For instance, Symbolic and Connectionist AI operate rhetorically using strong orientational metaphors – namely, the 'mind is a machine' and 'computers are like brains' respectively. At the heart of symbolic AI lays the project to translate mental experience into algorithmic formulations based on the assumption that mental operations are mechanistic, in that the mind can be seen as a machine. Connectionist AI foregrounds input-output models of information processing, and presumes that artificial neural nets are models that convey neurological realism, importing the viewpoint that computers are like brains. The 'mind is a machine' and 'computers are like brains' are orientational metaphors signifying conceptual frameworks that interpret mental experience through the generalizations of these concepts.

Richard Boyd explains that the 'mind is a computer' metaphor also gave rise to a number of generative concepts such as information-processing, encoding, decoding, feedback, memory stores among others that played an influential role in computational psychology:[37]

> The claim that thought is a kind of "information processing" and that the brain is a sort of computer.
>
> The suggestion that certain motoric or cognitive processes are "preprogrammed."

> Disputes over the issue of the existence of an internal "brain-language" in which "computations" are carried out.
> The suggestion that certain information is "encoded"or "indexed" in "memory store" by "labeling," whereas other information is "stored" in "images."
> Disputes about the extent to which developmental "stages" are produced by the maturation of new "preprogrammed" "subroutines," as opposed to the acquisition of learned "heuristic routines," or the development of greater "memory storage capacities" or better "information retrieval procedures."
> The view that learning is an adaptive response of a "self-organizing"machine.
> The view that consciousness is a "feedback" phenomenon.[38]

Boyd articulates that some metaphors are "theory constitutive" in that they play much more than an exegetical role and help construct some of the salient theoretical concepts in a particular framework, engendered by the principal metaphor:

> The prevalence of computer metaphors shows an important feature of contemporary theoretical psychology: a concern with exploring analogies, or similarities, between men and computational devices has been the most significant factor influencing postbehaviorist cognitive psychology... Moreover, it is clear that these computer metaphors are theory-constitutive: psychologists do not, generally speaking, now know how to offer literal paraphrases which express the same theoretical claims. This is made clearly evident by the current discussion among psychologists and philosophers about the doctrine called "functionalism"... It is widely agreed that some version or other of the doctrine that mental and psychological states are functional states of organisms represent the cognitive content of the metaphorical statement that the brain is a sort of computer. But even among psychologists and philosophers who are convinced that functionalism is true, there is profound disagreement about important issues regarding its interpretation. Thus, this metaphor and other computer metaphors employed in psychological theorizing, share with more typical interaction metaphors, at least for at time, the property that their cognitive content cannot be made explicit.[39]

Theory-constitutive metaphors can be seen in connectionist literature as well. Since it is assumed that computers are like brains, therefore neurobiological terms such as neural nets, synapses among others are used in descriptions of parallel processing computers.

David E. Rumelhart, one of the influential pioneers in Connectionist AI, discounts the overemphasized distinction between the literal and the figurative by pointing out the lack of a psychological correlate in the underlying process involved in apprehending this distinction.[40] Metaphors play a key role in language acquisition such that it is customary for children to use old words to apply for new situations:

> Normally speaking, the process of language comprehension and production for a young child not fully familiar with the conventional range of application of a term must proceed through a process of fitting the aspects of the current situation into the closest lexical concept already available. Often this will conform with the conventional application of the term and it will therefore appear that the child is using the bit of language "literally." Just as often, the child will apply the concepts in a nonstandard way and appear to generate "nonliteral" or "metaphorical" speech. Thus, for example, if the term "open" is learned in the context of (say) a child's mouth being open, and then it is applied to a door or a window, the child will appear merely to be demonstrating an understanding of the term. On the other hand, if the child uses the term "open" to mean "turn on" (as with a television set or a light) the child will be perceived as having produced a metaphor. Yet the process of applying words to situations is much the same in the two cases – namely that of finding the best word or concept to communicate the idea in mind. For the child the production of literal and nonliteral speech may involve *exactly* the same process.[41]

Although, Rumelhart concedes that there is a distinction between literal and figurative meanings – the point that is driven use that traditional conception of metaphoric descriptions as a special form of language use is perhaps overstated. Perhaps, this could be seen as a rhetorical justification for PDP models as indexes of human cognition in the face of criticism that connectionism has not displayed adequate evidence to warrant claims about neural realism.

A rhetorical understanding of the human-machine analog brings forth the attribution or mapping of each other's qualia onto the other practiced by theorists, who attribute mechanical reasoning as a defining cognitive quality for humans and by the same token, translate subjective qualia into algorithmic operations. The analogies are established on principles of "a) structural consistency (one-to-one correspondence), b) relational focus (relational systems are preserved and object descriptions disregarded), c) systematicity, d) no extraneous association (only commonalites strengthen an analogy), e) no mixed analogies and f) analogy is not causation."[42]

Michael J. Reddy states that metaphors are so pervasive that they engulf all aspects of human thought.[43] Reddy postulates the notion of a 'conduit'metaphor which implies that linguistic expressions are containers of meaning:

> The logic of the framework we are considering – a logic which will henceforth be called the *conduit* metaphor – would now lead us to the bizarre assertion that words have "insides" and "outsides" After all, if thoughts can be "inserted," there must be a space "inside" wherein the meaning can reside. But surely the English language, whatever metaphysical meanderings it may have been guilty of thus far, cannot have involved us in this kind of patent nonsense. Well, a moment's reflection should nudge anyone into remembering that "content" is a term used almost synonymously with "ideas" and "meaning." And that recollection is quite meaning-full (sic) in the present context. Numerous expressions make it clear that English does view words as containing or failing to contain thoughts, depending on the success or failure of the speaker's "insertion" process.[44]

Expressions such as "that *thought is in* practically every other *word*," "the *sentence was filled with emotion*," "the *lines* may rhyme, but *they are empty* of both *meaning* and *feeling,* " "your *words are hollow* – you don't mean them" among others are common examples of conduit metaphors found in every day expression.[45] How conduit metaphors apply in connectionist AI is manifest in the inside-outside dichotomy in the description of processing units with terms such as input, hidden and output respectively. Cognition takes places in an inner space of interconnected neural nets (inside), acted on by input units from an external environment (outside). Symbolic

AI also displays elements of the conduit metaphor, in that, thought is supposed to be 'contained within' symbol-manipulation. The mind is seen as a finite, problem space in which a large number of algorithmic procedures or mental operations are performed that give rise to what we consider 'intelligence.' There are programmed routines and subroutines within the mind that enables the extrapolation of the machine analogy to humans, and by the same token, human mental functions to machines. Therefore, metaphors in these cases are seen much more than mere comparisons.

Andrew Ortony goes beyond the substitutional or comparison view by articulating the following:

> It is often claimed that metaphors are merely implicit comparisons to be contrasted with similes, which are explicit ones. I have very little faith in the view: first, because I do not think that is true of all metaphors; and second, because even if it were, it would be totally unilluminating. The fact that metaphors are frequently used to make comparisons, if it is a fact, does not mean that metaphors *are* comparisons. A metaphor is a kind of *use* of language, whereas a comparison is a kind of psychological process, which although quite possibly an essential component of certain kinds of language use, is not the same thing as such a use.[46]

Implicit in Ortony's claims is the suggestion that a mere comparison view of metaphors reduce them to analogies, rendering them dispensable to the projected descriptions of a process. Applying this concept to our analysis, the manner in which computational psychologists borrow terms from computers is not only to establish analogies, but also to metonymically explain one in terms of another.

Max Black articulates the need to discriminate two aspects, which he refers to as "emphasis" and "resonance" while dealing with metaphorical descriptions:

> Given an active metaphorical statement, it would be useful to discriminate two aspects, which I shall call emphasis and resonance. A metaphorical utterance is

> *emphatic*, in my intended sense, to the degree that the producer will allow no variation upon or substitute for the words used…Plausible opposites to "emphatic" might include: "expendable," "optional," "decorative," and "ornamental." (Relatively dispensable metaphors are often no more than literary or rhetorical flourishes that deserve no more serious attention than musical grace notes). Emphatic metaphors are intended to be dwelt upon for the sake of their unstated implications. Their producers need the receiver's cooperative in perceiving what lies *behind* the words used. How far such interpretative response can reach will depend upon the complexity and power of the metaphor-theme in question: Some metaphors, even famous ones, barely lend themselves to implicative elaboration, while others, perhaps less interesting, prove relatively rich in background implications. For want of a better label, I shall call metaphorical utterances that support a high degree of implicative elaboration *resonant*.[47]

From the symbolic AI camp, computational psychologists find the computer to be an indispensable and emphatic model of cognition. The indispensability of the computer model is a strong metaphor reflected that the conflation of distinctions is so pervasive, that the metaphoric dimensions gradually slips away into the consciousness:

> The view that cognition can be understood as computation is ubiquitous in modern cognitive theories, even among those who do not use computer programs to express models of cognitive processes. One of the basic assumptions behind this approach, sometimes referred to as information processing, is that cognitive processes can be understood in terms of formal operations carried out on symbol structures.[48]

Some clarification is necessary here. If the computer is to be a strong metaphor for cognition, its usage should be inconspicuous in a sense that it deflects attention from itself. Its usage should be subtle enough, that the focus is not on the metaphor per se, but on the descriptions generated through the model that makes the conflation between the subject and the metaphoric predicate possible. Computational psychology downplays the comparison between "humans" and "machines" because the tendency to look for plain similarities of effects is always accompanied by the concomitant search for dissimilarities. Instead, computational psychologists affirm the structural isomorphism between humans and machines at face-value, to such an extent, that questions to the contrary will usually be explained away as a problems of unspecified complexity, that will be eventually resolved with sufficient parsing.

The computer metaphor is not only considered indispensable but also deflects attention from itself; since in computational psychology, more attention is paid to cognitive processes characterized by computation, thereby rendering computation and cognitive processes interchangeable. Therefore, when Turing replaces the question 'can machines think?' with 'can digital computers engage in intelligent behavior that would require intelligence if performed by humans?' the emphasis is placed less on machine intelligence on its own terms and more on intelligent behavior.

As stated earlier, the strength of a metaphor lies in its ability to be co-opted into a framework of scientific realism, such that the differences are so completely eroded in the minds of those who propagate that specific orientation. The manner in which metaphoricity has lapsed into literality such that the metaphor is completely forgotten is echoed in the words of computational psychologist Zenon W. Pylyshyn who asserts, "computation is a literal rather than metaphoric view of cognition."[49] Pylyshyn critiques the linguistic opacity of metaphoric descriptions due to their purported lack of proper reference; therefore he suggests the following:

> My own tentative feeling is that the difference between literal and metaphorical description lies primarily in such pragmatic considerations as (1) the stability, referential specificity, and general acceptance of terms; and (2) the perception, shared by those who use the terms, that the resulting descriptions characterize the world *as it really is,* rather than being a convenient way of talking about it, or a way of capturing superficial resemblances.[50]

The lack of a coherent demarcation between literal and metaphorical description and the lack of specification about the referential nature of particular words, by default has created a wrong-headed blurring of distinctions. The only caveat that I might insert is that it is not clear how referential specificity or a general acceptance of terms by themselves can *ex cathedra* eliminate the metaphoric dimensions of a term. It is not clear how metaphors are precluded from possessing the elements of referential specificity even if one were to embrace a scientific realist view of language. Also, even apparently non-metaphoric terms such as stability and referentiality are at some level, topographic terms that are imported into the realm of ideas not unlike the manner in which terms such as 'depth,' 'breadth' and 'shallowness' are used to assess the treatment of ideas. Pylyshyn's distinction between literal and conveyed meanings lies in the realm of *seeing* (visual metaphor) the world *as it is,* rests on unmediated, nonconstructivist assumptions.

Metaphors as Terministic Screens

The role of metaphors as terministic screens and how the screens play out will be elaborated upon in chapters 4, 5 and 6 respectively. Symbolic AI begins with the computer metaphor and we see a shift from a mechanistic metaphor to a biological one with the emergence of connectionism. Instead of using the machine (computer) as a model of cognition, connectionists consider the parallel processing brain as an exemplar for building artificial neural networks. But the brain is considered a machine – therefore, the discourse of AI comes full circle.

Machine		Human Cognition
A	$\rightarrow$	B
A_1 (Digital computer)	$\rightarrow$	B_1 (Human cognition according to Symbolic AI)
A_2 (Artificial Net)	$\leftarrow$	B_2 (Human cognition according to Connectionist AI)

From the above-mentioned equation, AI is fundamentally based on the assumption that machines provide a good model for human cognition. With symbolic AI, the idea of a digital computer is mapped onto accounts of human cognition. With connectionist AI, the model of human cognition provided by the brain is mapped onto artificial nets. Connectionist AI moves in the reverse direction compared to symbolic AI. But artificial nets and the neurophysiological brain are ultimately seen as machines. Therefore, the machine metaphor remains the overarching lens with regard to human cognition. This metaphor dominates much of computational psychology, becoming both an avenue of further exploration using algorithms as indices of thought and also becoming a gatekeeper or perhaps a circumscribing influence on other types of discourse.

The idea behind metaphors as terministic screens purports to the notion that much of the discourse revolving around Artificial Intelligence deals with representing mental activity, establishing the processes by which mental activity comes about, using a specialized linguistic taxonomy borrowed from computer science to describe human psychology and by the same token using humanistic taxonomy to describe machine interaction. The very act of using a particular nomenclature to represent a nomological process, in itself, becomes a terministic screen that selects particular epistemological orientations, while deflecting competing orientations. To conclude, Prelli's remarks conveys a sense of how the process of selection and deflection are rhetorically, inextricably intertwined in scientific rhetoric reaffirming the Burkean notion that our screens accompany us everywhere:

> Despite the normative goal of reducing ambiguities, scientific rhetors still have to choose discursive strategies that will convince colleagues that specific aims and claims do further comprehension, and should be incorporated into the field's evolving literature as reasonable. These strategies may, of course, be selected consciously or according to habit originally inculcated by training in scientific methods. In either case, the overarching norm of all scientific persuasion reveals relevant sayables and strategies that make logical sense. Put in rhetorical terms, the normative principle governing scientific logic directs attention toward certain *topoi* and away from others.[51]

These screens are ultimately reflective of not only our constructions of cognition, but also of our constructions of human nature itself. However, regardless of one's philosophical framework, AI forces us to ask the question 'are we unique?' The lack of consensus among serious scholars of mind, and the subsequent efforts at justification makes AI a fertile ground for scholars in argumentation.

Summary

The sense in which metaphors will be employed in the following chapters pertains to the notion that metaphors are generative tools. The generation of concepts takes place through the mapping, and not so much through the metaphor by itself. When the metaphor employed is appropriate and seen as a lens, the metaphor generates a conceptual mapping of vocabularies from one entity onto another. In symbolic AI, the machine metaphor generates a conceptual mapping of computational vocabulary onto human cognition and with connectionist AI, the mapping starts with the brain as the source domain. This will be explored in greater detail in the analyses chapters.

Notes

[1] Alan G. Gross, *The Rhetoric of Science* (Cambridge, Massachusetts: Harvard University Press, 1996), 81.
[2] Gross, *The Rhetoric of Science*, vi-vii.
[3] Max Perutz, "Letter," *Science* 164: 1537-1538.
[4] John Lyne and Henry F. Howe, "Punctuated Equilibria: Rhetorical Dynamics of a scientific controversy," in *Landmark Essays on Rhetoric of Science Case Studies*, ed. Randy Allen Harris (New Jersey: Lawrence Erlbaum, 1997), 69.
[5] Albert R. Jonsen and Stephen Toulmin, *The Abuse of Casuistry: A History of Moral Reasoning* (Berkeley: University of California Press, 1988), 8.
[6] Jonsen and Toulmin, *The Abuse of Casuistry*, 306-307.
[7] CH. Perelman and L. Olbrechts-Tyteca, *The New Rhetoric: A Treatise on Argumentation*. Trans. John Wilkinson and Purcell Weaver (Notre Dame: University of Notre Dame Press, 1969), 14
[8] Perelman and Olbrechts-Tyteca, *The New Rhetoric*, 385.
[9] Richard McKeon, *On Knowing – The Natural Sciences*. Lectures compiled and edited by David B. Owen and Zahava McKeon (Chicago: The University of Chicago Press, 1994), 167.
[10] Jürgen Habermas, *Knowledge and Human Interests*, trans. J. J. Shapiro (Boston: Beacon Press, 1971), 309.
[11] Habermas, *Knowledge and Human Interests*, 311, 195.
[12] Charles Arthur Willard, *Argumentation and the Social Grounds of Knowledge* (Alabama: The University of Alabama Press, 1983), 136.
[13] Richard B. Miller, *Casuistry and Modern Ethics* (Chicago: The University of Chicago Press, 1996), 5.
[14] Perelman and Olbrechts-Tyteca, *The New Rhetoric*, 378.
[15] Richard Weaver, *The Ethics of Rhetoric* (La Salle, Indiana: Gateway, 1953), 56-57.
[16] Richard Weaver, *Language is Sermonic*, ed. Richard L. Johannesen, Rennard Strickland and Ralph T. Eubanks (Baton Rouge: Louisiana State University Press, 1970), 45.
[17] Allen Newell and Herbert A. Simon, "Computer Science as Empirical Inquiry: Symbols and Search," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: MIT Press, 1997).
[18] Newell and Simon, "Computer Science as Empirical Inquiry," 94.
[19] Herbert A. Simon, "Machine as Mind," in *Machines and Thought: The Legacy of Alan Turing Vol I*, ed. P. J. R. Millican and A. Clark (Oxford: Clarendon Press, 1996), 91.
[20] Newton Garver and Seung-Chong Lee, *Derrida & Wittgenstein* (Philadelphia: Temple University Press, 1994), 67.
[21] Paul Churchland, "Some Reductive Strategies in Cognitive Neurobiology," *Mind* XCV, 379: 279-309.
[22] J. Greene, "Evaluative cognitive explanations of communicative phenomena," *Quarterly Journal of Speech*, 70: 241-254
[23] Robert D. McPhee, "Cognitive Perspectives on Communication: Interpretive and Critical Responses," in *The Cognitive bases of Interpersonal Communication*, ed. Dean E. Hewes (Hillsdale, NJ: Lawrence Erlbaum, 1995), 225-245.
[24] David E. Rumelhart, James L. McClelland and others, *Parallel Distributed Processing: Exploration in the Microstructure of Cognition* (Cambridge, Massachusetts: MIT Press, A Bradford Book, 1986), 10.
[25] Andy Clark, *Mindware: An Introduction to the Philosophy of Cognitive Science* (New York: Oxford University Press, 2001).

[26] Philip E. Agre, *Computation and Human Experience* (Cambridge: Cambridge University Press, 1997), 86.
[27] Agre, *Computation and Human Experience*, 86.
[28] Richard Boyd, "Metaphor and Theory Change: What is 'Metaphor' a Metaphor for?" in *Metaphor and Thought*, ed. Andrew Ortony (Cambridge: Cambridge University Press, 1979), 369.
[29] John Locke, *An Essay Concerning Human Understanding*, ed. J. W. Jolton (London: Dent, 1961), 89.
[30] Locke, *An Essay Concerning Human Understanding*, 89.
[31] Locke, *An Essay Concerning Human Understanding*, 129.
[32] Locke, *An Essay Concerning Human Understanding*, 129.
[33] David Hume, *A Treatise of Human Nature* (Oxford: Clarendon Press, 1965), 253.
[34] George Lakoff and Mark Johnson, *Metaphors We Live By* (Chicago: The University of Chicago Press, 1980), 3.
[35] Lakoff and Johnson, *Metaphors We Live By*, 10.
[36] Lakoff and Johnson, *Metaphors We Live By*, 14.
[37] Richard Boyd, "Metaphor and theory change: What is "metaphor" a metaphor for?," in *Metaphor and Thought, Second Edition*, ed. Andrew Ortony (Cambridge: Cambridge University Press, 1993), 481-532.
[38] Boyd, "Metaphor and theory change," 486.
[39] Boyd, "Metaphor and theory change," 487.
[40] David E. Rumelhart, "Some problems with the notion of literal meanings," in *Metaphor and Thought, Second Edition*, ed. Andrew Ortony (Cambridge: Cambridge University Press, 1993), 71-82.
[41] Rumelhart, "Some problems with the notion of literal meanings," 73.
[42] Dedre Gentner and Michael Jeziorski, "From metaphor to analogy in science," in *Metaphor and Thought, Second Edition*, ed. Andrew Ortony (Cambridge: Cambridge University Press, 1993), 447–480 (450).
[43] Michael J. Reddy, "The Conduit Metaphor: A case of frame conflict in our language about language," in *Metaphor and Thought, Second Edition*, ed. Andrew Ortony (Cambridge: Cambridge University Press, 1993), 164-201.
[44] Reddy, "The Conduit Metaphor," 168.
[45] Reddy, "The Conduit Metaphor," 168.
[46] Andrew Ortony, "Similarity in similes and metaphors," in *Metaphor and Thought, Second Edition*, ed. Andrew Ortony (Cambridge: Cambridge University Press, 1993), 342-356 (344).
[47] Max Black, "More about Metaphor," in *Metaphor and Thought, Second Edition*, ed. Andrew Ortony (Cambridge: Cambridge University Press, 1993), 19-41 (25-26).
[48] Zenon Pylyshyn, "Computation and Cognition: Issues in the Foundations of Cognitive Science," *Behavioral and Brain Sciences* 3 (1980): 111-169 (111).
[49] Zenon W. Pylyshyn, "Metaphorical imprecision and the "top-down" research strategy," in *Metaphor and Thought, Second Edition*, ed. Andrew Ortony (Cambridge: Cambridge University Press, 1993), 543-558 (553).
[50] Pylyshyn, "Metaphorical Imprecision and the "top-down" research strategy," 556.
[51] Prelli, *A Rhetoric of Science*, 125.

CHAPTER III

ANTHROPOMORPHIZING

> For, as human beings, we must inevitably see the universe from a center lying within ourselves and speak about it in terms of a human language shaped by the exigencies of human intercourse. Any attempt rigorously to eliminate our human perspective from our picture of the world must lead to absurdity.[1]

The elimination of human perspective from human accounts of the universe is impossible since even such an act of negation involves a conscious decision by agents to get rid of our lenses. Human understanding, at the risk of sounding tautological, will always possess elements of the 'human' which manifests itself in personal predispositions, training, methodological practices, philosophical frameworks, linguistic habits among others. The 'elements of the human' may be tentatively referred to as 'anthropomorphizing' (note to committee members: I will change this term as soon as I find a better one). Some clarification is necessary in order to avoid confusion with a closely related word namely 'anthropomorphism.' The latter refers to the projection of human qualia onto animals or artifacts. In other words, one is referring to the attribution of 'likeness.' Anthropomorphizing, on the other hand, refers to the *intentions* or *motives* behind the conscious act of interpreting the 'world' through the active participation of human agents who use specific linguistic lenses (that makes most sense to them – either through their training, beliefs or so forth). If metaphors as terministic screens provide us lenses with which to see the world, anthropomorphizing pertains to the intentions or rhetorical motivation behind.

Anthropomorphizing refers to the intentional articulation of sense-making linguistic habits or mechanisms. For instance, if one were to say that 'thought is computation' and that 'all thinking hinges on some type of computation' – the anthropomorphizing lies in the desire to understand the mind primarily in terms of a rationalist paradigm, of which mathematical logic or programming is the most accurate expression. The above-mentioned statement is unmistakably true of symbolic AI. With connectionist AI, anthropomorphizing lies in the desire to understand all mental activity in terms of an empirical/neurological paradigm, of which neural nets are the most accurate expression.

Anthropomorphizing, in the context of this discussion, should be seen as the motivation, namely the desire to understand the machine in terms of the human – and indirectly, understanding the human in terms of the machine by making the human a machine. The purpose of this chapter is to discuss anthropomorphizing as the rhetorical motivation in AI's discourse of mind. The argument that I will present is tedious and needs elaboration, in part because scholarly communities have not often considered anthropomorphizing to be a part and parcel of scholarly discourse. For a variety of historical and intellectual reasons, 'intentionality' and 'agency' closely tied to anthropomorphizing have been eclipsed from many intellectual discussions – therefore, the path to reinstating 'personal' or 'anthropomorphized' knowledge will be a circuitous one at first. Therefore, the spirit of this chapter is not to engage in reactionary, intellectual combat against the fruitful and productive accomplishments of earlier theories, but only to re-invite reflection on the ineliminability of perspective manifest in our discourse being shaped by our preferences and experiences, and by the same token how the discourse reshapes our perceptions of who we are vis-à-vis our status as participant observers.

From Spectator to Participant Observer

At least two narratives vis-à-vis the relationship between human agency and the apprehension of scientific knowledge are in popular circulation. One story describes a curious spectator, a discoverer or reporter who observes the unfolding cosmic drama of events from a particular vantage point in time, describing and evaluating both macro and microcosmic phenomena. The spectator is sometimes typecast, for lack of a better term, an imprimatur or is part of a committee of fellow imprimaturs settling on what is and what is not worthy of study, specifying the conditions under which a study may be conducted and invoking the authority of an accepted theory or paradigm (until it changes) as a framework to justify his or her assertions. The spectator may tentatively be granted the freedom to have minimal intellectual sympathies in the form of hunches, conjectures or hypotheses as long as these sympathies in no way affects or skews what is being investigated, and at the same time these sympathies should either be confirmed or disconfirmed by rigorous scientific experimentation.

The other view portrays the human agent as an active participant in the process of apprehending knowledge, thus equating the process of participation as a form of construction as well. In simpler words, our lenses are our faithful companions everywhere we go not only influencing our perception but also shaping the 'world' we inhabit; granted the possibility that our lenses may be replaced with newer ones with the passage of time and the arrival of new knowledge about a particular subject.[2] Michael Polanyi articulates what he considers to be the domain of personal knowledge, that engulfs the sciences as well:

> It is by his assimilation of the framework of science that the scientist makes sense of experience. This making sense of experience is a skilful act which impresses the personal participation of the scientist on the resultant knowledge. It includes the skill of carrying out correctly the measurements which verify scientific predictions or the observations by which scientific classifications are applied...The tracing of personal knowledge to its roots in the subsidiary awareness of our body as merged in our focal awareness of external objects, reveal not only the logical structure of personal knowledge but also its dynamic sources...In this sense, I should say that an object is transformed into a tool by a purposive effort envisaging an operational field in respect of which the object guided by our efforts shall function as an extension of our body....Like the tool, the sign or the symbol can be conceived as such only in the eyes of a person who *relies on them* to achieve or to signify something. This reliance is a personal commitment which is involved in all acts of intelligence by which we integrate some things subsidiarily to the center of our focal attention. Every act of personal assimilation by which we make a thing a form of extension of ourselves through our subsidiary awareness of it, is a commitment of ourselves; a manner of disposing of ourselves.[3]

The idea of personal knowledge entails a commitment to a particular way of viewing the world and rests on the premise that in cases where an individual is not completely knowledgeable about a subject matter, he or she can always fall back on the authority of the scientific community. Thus personal knowledge becomes a form of communal knowledge, given the fact that a single individual can only comprehend or apprehend so much:

> Any attempt to define the body of science more closely comes up against the fact that the knowledge comprised by science is not known to any single person. Indeed, nobody knows more than a tiny fragment of science well enough to judge its validity and value at first hand. For the rest he has to rely on views accepted at second hand on the authority of a community of people

> accredited as scientists. But this accrediting depends in turn on a complex organization. For each member of the community can judge at first hand only a small number of his fellow members, and yet eventually each is accredited by all. For each member of the community can judge at first hand only a small number of his fellow members, and yet eventually each is accredited by all. What happens is that each recognizes as scientists a number of others by whom he is recognized as such in return, and these relations form chains which transmit these mutual recognitions at second through the whole community. This is how each member becomes directly or indirectly accredited by all. The system extends into the past. Its members recognize the same set of persons as their masters and derive from this allegiance a common tradition, of which each carries on a particular strand.[4]

The idea of comprehension whether personal or communal, involves the mental faculties that engages in the act of measuring; and measurements do take place in a temporal context. The elimination of temporality or mental faculties from the realm of knowing is inconceivable, and such an act in turn entails a recognition of some type of agency.

St. Augustine illustrates this point well commenting on how the observations of the agent are ultimately participatory and not merely tangential:

> Let us consider the case of a bodily voice. The voice begins to sound, it continues to sound, and then it stops sounding. Now there is silence; the voice is past and is no longer a voice. Before it began to sound, it was in the future and could not be measured because it did not exist, and now it cannot be measured because it no longer exists...What, then, is it that I measure? Where is that short syllable by which I measure? Where is that long syllable which I measure? Both have sounded, have fled away, have gone into the past, and no longer exist; and yet I do measure...Therefore, what I do measure are not the syllables themselves (they no longer exist) but something in my memory which remains there fixed.[5]

The point he makes is undeniable, in that even today the idea of a measurement is inconceivable without the intervention of the mind although the tools of measurement or analysis are much more than sophisticated than our predecessors. St.

Augustine may also be forgiven for his lack of specification and lack of adequate neuroscientific evidence in asserting the following:

> It is in you, my mind, that I measure time. Do not interrupt me, or rather, do not allow yourself to be interrupted by the thronging of your impressions. It is in you, I say, that I measure time. As things pass by they leave their impression in you; this impression remains after things have gone into the past, and it is this impression which I measure in the present, not the things which, in their passage, caused the impression. It is this impression which I measure when I measure time. Therefore, either this itself is time or else I do not measure time at all.[6]

To make a broad sweep, grappling with 'measurements' or 'time' or just about any concept worthy of intellectual inquiry cannot be negotiated without some sense of participatory agency. The participatory dimension is what makes knowledge personal. The personal dimension should not be confused for solipsistic knowledge or even an ersatz science that is a quaint throwback to an anthropocentric, Ptolemaic cosmology. Planetary motion or the speed of light is by no means dependent on human perception. But in order to understand the idea of planetary motion or any other natural phenomena, one has to seek recourse to a language or a vocabulary that is amenable to the discursive community such that it is precisely the selection of the linguistic vocabulary that makes the human agent a participant and not so much that the agent influenced or wished into existence the laws of planetary motion. However, the selection of a particular vocabulary may influence the inquiry, either by shedding more insight or by blinding the investigator to other aspects. Therefore the very selection of metaphors and conceptual vocabularies is an act of anthropomorphizing, meaning that we try to understand the external world in a human language that enhances or even defers comprehension.

An example briefly worth mentioning is the scientific controversy between proponents of the big bang theory and advocates of the short-lived steady state theory. Georges Lemaitres whose ideas of an atomic universe sowed the seeds for the later big bang theory of American Physicist George Gamov suggested that the beginning of the universe in the mind of the common folk is always interconnected with notions of space and time, however, he argues that in the atomic process 'space' and 'time' are no more than statistical entities:

> If the world had begun with a simple quantum, the notion of space and time would altogether fail to have meaning at the beginning; they would only begin to have a sensible meaning when the original quantum has been subdivided into a sufficient number of quanta. If this suggestion is correct, the beginning of the world happened a little before the beginning of space and time. I think that such a beginning of the world is far enough from the present order of nature to be not at all repugnant...we could conceive the beginning of the universe in the form of a unique atom, the atomic weight of which is the total mass of the universe. This highly unstable atom would divide in smaller and smaller atoms by a kind of super-radioactive process.[7]

Although the idea of "a little before" the beginning of space and time seems contradictory, but Lemaitres' atomic cosmology was further refined and developed by nuclear physicist George Gamov who speculated about the origins of the universe in terms of the 'big bang theory' where the early universe was seen as a "nuclear oven in which the elements constituting our present universe was once cooked."[8] Gamov and his collaborators' contributions were unique in that he united the world of astrophysics with nuclear physics.[9] The language of nuclear physics inspired the cosmology of the big bang theory and in the popular imagination it was not difficult to conceive a cosmology derived from nuclear physics i.e. if a destructive process can be effected through nuclear weapons the same could apply for a constructive process, namely a big-bang cosmology based on a neutron-capture process. It was suggested that when protons collide with lighter elements there is a possibility that the neutrons ejected from the nuclei of these lighter elements may glue themselves to the nuclei of heavier elements, thereby creating a possibility of the creation or formation of even heavier nuclei. The notion of a chain of consecutive neutron captures is said to have developed the universe we inhabit.

The rival paradigm to Gamov's big bang theory is Fred Hoyle's "Steady State Theory."[10] Fred Hoyle's personal view of cosmology conceives the nature of a scientific hypothesis as such:

> It is often held that scientific hypotheses are constructed, and are to be constructed only after a detailed weighing of all possible evidence bearing on the matter, and that then and only then may one consider, and still only tentatively, any hypothesis. This traditional view, however, is largely incorrect, for not only is it absurdly impossible of application, but it is contradicted by the history of the development of any scientific theory. What happens in practice is that by intuitive insight, or any inexplicable inspiration, the theorist decides that certain features seem to him more important than others and capable of explanation by certain hypotheses. Then basing his study on these hypotheses the attempt is made to deduce their consequences. The successful pioneer of theoretical sciences is he whose intuitions yield hypotheses on which satisfactory theories can be built, and conversely for the unsuccessful (as judged from a purely scientific standpoint).[11]

Hoyle appears to acknowledge the role of the scientist's intuition and hunches in creating hypotheses and subsequently theories. Hoyle's theory is no longer considered in most circles as a viable candidate for cosmological origins, yet from a rhetorical perspective, one of his grouses against the big bang hypothesis is worth noting in that he highlights the inaccessibility to direct observation as one of the chief failings of this alternative cosmology: "On scientific grounds this big bang work is much the less palatable of the two. For it is an irrational process that cannot be described in scientific terms…since it puts the basic assumption out of sight where it can never be challenged by a direct appeal to observation."[12]

Advocates of Gamov's theory on the other hand invoke the authority of Hans Albrecht Bethe for his pioneering work in explaining the energy production of the sun through thermo-nuclear reactions. This is closely tied with the work at Caltech's Kellogg laboratory where nuclear physicists were studying the reactions between protons and carbon nuclei.[13] Nobel laureate William Fowler in a tribute to Bethe's contributions remarked: "Bethe's paper told us that we were studying in the laboratory processes which are occurring in the sun and other stars. It made a lasting impression on us."[14] Although the controversy regarding Gamov and Hoyle can make a fascinating study for those interested in scientific argumentation from a rhetorical point of view, the purpose of bringing in this analogy is to show how individual lenses and orientations of certain physicists have been instrumental in shaping new disciplines. Such an approach would by no means undermine the presence of an external world or drift into a solipsistic mentalism, but will only shed light on how the choice of scientific language in itself is epistemology-generating. Due to lack of space, this discussion will not explore the metaphoric delineation of these two theories, - but what can be taken away from this debate, - is that even scientific knowledge is personal, and that personal knowledge is a matter of anthropomorphizing. The reader should not attach too much credence to the view that personal knowledge would degenerate into 'anything goes' discourse because after all, we are dealing with well-established theories that have been legitimized by respective scientific communities. Therefore, 'personal' should be seen as '-the person in-' the discourse as opposed to an impersonal model of discourse where agents have nothing to do with the way in which we interpret the world.[15]

Anthropomorphizing is Personal Knowledge

The idea behind anthropomorphizing helps us understand at least two-things: a) it underscores the intentionality of human agents in constructing theories, even in cases where these paradigms appear to be normative and ontological descriptions of an external reality independent of the observer (cases such as the origins of the universe may fall under this realm), b) it demonstrates how even theories that seem to have 'lives of their own' (in explaining the natural or social world) with an illusion of autonomy are by themselves linguistic and rhetorical impositions of human subject trying to understand the world in terms of orientations most amenable.

Taking such a position might invite criticism that would preemptively conflate 'anthropomorphizing' with some kind of neo-Ptolemaic, geocentric cosmology in which humans are accorded centrality. However, without necessarily denying a material and external reality distinct from the 'knower,' one can still take a middle ground by eschewing a simon-pure and definitive dissociation between the knower and the known, largely because the knower has to undeniably rely on his or her own cognitive lenses if he or she were to make sense of the world. In addition, these lenses affect the way in which we see the world because they are mediated through language. Making such a claim does not mean that the observer is one with the 'rock,' 'planet' or the particular phenomenon being observed. Instead it only rejects the idea that our experiential frames do not come into play while studying particular phenomena. The same could be said if one were to replace one human agent with another.

Polanyi argues that the dissociation of science from personal knowledge is a common refrain in intellectual circles:

> The prevailing conception of science, based on the disjunction of subjectivity and objectivity, seeks – and must seek at all costs – to eliminate from science such passionate, personal, human appraisals of theories, or at least to minimize their function to that of a negligible by-play. For modern man has set up as the ideal of knowledge the conception of natural science as a set of statements which is 'objective' in the sense that its substance is entirely determined by observation, while its presentation may be shaped by convention. This conception, stemming from a craving rooted in the very depths of our culture, would be shattered if the intuition of rationality in nature is to be acknowledged as a justifiable and indeed essential part of scientific theory. That is why scientific theory is represented as a mere economical description of facts; or as embodying a conventional policy for drawing empirical inferences; or as a working hypothesis, suited to man's practical convenience – interpretations that all deliberately overlook the rational core of science.[16]

What Polanyi describes as the 'rational core' of science is ultimately a rational motivation for a rational orientation. The need for a systematic and logical understanding of the world not only entails a linguistic/logical imposition of a framework, but also entails an ordering of entities in a verifiable fashion that makes the discourse possible. As Albert Einstein explains, "concepts which have proved useful for ordering things easily assume so great an authority over us that we forget their terrestrial origin and accept them as unalterable facts."[17] The presuppositions of rationality that are quite necessary in the sciences have terrestrial or human origins, in that we try to make sense of the cosmos in a language that personally enhances understanding. In the archaeology of ideas, the search for order or meaningful patterns in nature and the subsequent selection of terms that amplify the visions of order are common. The image of order is not just a mental abstraction, but also a concrete motivation expressing itself manifestly through the theories and hypotheses generated and sustained over a period of time. An emphasis on 'anthropomorphizing' is by no means tautological in that it plays out through the use of metaphors, metonymies and synecdoches, among other rhetorical device that are wittingly or unwittingly employed by scientists in generating their discourse.

The symbolicity of the world we inhabit permeates every human realm. As Ludwig von Bertalanffy explicates:

> Except for the immediate gratification of biological needs, man lives in a world not of things, but of symbols. A coin is a symbol for a certain amount of work performed or of food and other available products. A written document is a symbol of past events. A word or a thought is a symbol of past events. A word or a thought is a symbol for an object or a relation. A book is a fantastic mass of symbols piled on one another. What determines human behavior beyond food and sex are symbolic needs – whether they signify social position, wealth, satisfaction of personal interests, or even scientific and artistic activities and the realization of high ethical standards.[18]

Linguistic signs are symbolic means of apprehending the 'real-world'; 'symbolicity' expresses itself in the form of 'metaphoricity' where one sign is appropriated in another context. 'Metaphoricity' lies at the core of human thinking in that gives a concrete object to work with; a frame of comparison; a frame of understanding of understanding an entity in terms of a similitude especially when present understanding is not adequate to grasp the term on its own terms. It is impossible to rid human language of metaphors, and this encompasses the sciences as well. 'Metaphoricity' may not be apparent when it comes to the level of specification; however, it is inevitable at the level of general theorization. Many metaphors can be found at the 'foundations' of the sciences or general intellectual inquiry as well – e.g. light is both a *wave* and a *particle*, 'selection' in evolutionary biology, the idea of a 'framework' as a framework, the idea of an 'idea,' the need for 'grounds,' the need for 'depth' in analysis, the need for 'breadth' in knowledge, the need to avoid 'shallowness.

The previous chapter argued that metaphors are the wheels that drive the theories about mind design in AI. If metaphors constitute the 'generative' devices and terministic screens, anthropomorphizing pertains to the motivation behind the selection of metaphors. Anthropomorphizing as a motivation, explains the human need to understand the universe in terms of a language that is constructed by us, the construction of which is simultaneously a selection and a deflection.

Anthropomorphizing in AI

Anthropomorphizing in symbolic AI takes the form of understanding cognition through the lens of mathematical logic or programming. The physical symbol hypothesis assumes that all thinking is symbolic. Symbol manipulation forms the basis for human intelligent action. Humans not unlike machines are considered symbol manipulators. Therefore, it became possible to import from the digital computer a whole arsenal of computational vocabulary facilitating the birth of computational psychology.

The anthropomorphizing that occurs may seem indirect, in that one starts with the assumption that machines are models for human cognition; therefore an extrapolation of the symbol manipulation into the realm of human thought is not farfetched. However, the anthropomorphizing is much more direct when one takes into account the consideration that humans are seen as machines, and that the mathematico-rationalist paradigm constitutes the lens with which we see the world.

The philosophy of symbolic AI, when seen from a sociological or rhetorical standpoint, conceives humans as automatons. The anthropomorphizing occurs in the self-automating motivation of humans to understand themselves in terms of machinery whose processes are self-regulating and predictable as best. Mechanical models of mind based on serial computing dominated the field of computational psychology, owing its inspiration to symbolic AI.

The case for anthropomorphizing is more direct with connectionism. Connectionism starts or claims to start with the human brain and use the human brain as an index for building parallel processing networks since the brain in itself is considered a parallel processing network. Connectionists consider the biological brain to be isomorphic with the digital brain. The connectionists invoke the ascendancy of the neuroscientific accounts of consciousness and the concomitant eclipse of purely speculative or folk psychological accounts of the mind. Their emphasis is on the range of intelligent behavior that is translatable, subsequently made possible by studying neural interaction in the brain. The empirical status ascribed to studies of mental activity gives it the sort of respectability that fields like physics, chemistry and biology have long enjoyed. The cross-fertilization of ideas across the neurosciences and physics especially makes cognitive psychology a thriving, empirical science. The metaphors that cognitive scientists use to characterize humans vis-à-vis machines are eloquent testaments of anthropomorphism where machines are rendered in humanlike form and the corollary where humans are regarded as biological machines at the neurological level.

In effect, both symbolic and connectionist AI can be treated as computational paradigms of mind. From a rhetorical perspective, these paradigms are insertions of anthropomorphic frames into the object of investigation. A paradigm, in itself, serves, as a governing lens, emerging from human *praxis* without which, as Thomas Kuhn suggests, the structured universe of scientific discourse is impossible.[19] Under the physicalist/materialist paradigm of mind design, consciousness is either reduced to phenomena that can be explained through sentential epistemologies (symbolic AI) or explained away or construed purely as a product of neurological activity (connectionism).

Part of the anthropomorphizing in AI is connected to a deep-seated, pervasive scientific determinism much more sophisticated than versions of nineteenth century determinism. Biological determinism turns into mechanistic or neurological determinism. Every action is a consequence of initial conditions outside the perimeter of our control by way of causal laws that are also outside our control. Underlying this view is a picture of the universe as a gargantuan machine, where at both the macroscopic and microscopic levels, all events, including human actions are determined by preceding events and causal laws. At a given time t, agent A when acted upon by conditions C will perform action a. A soft determinist might even attempt to reconcile human freedom, in a limited sense with causal determinism (in order to foster an atmosphere of peaceful coexistence between competing ideologies or plainly for exercising ethical responsibility for one's actions). The soft determinist may argue that agent *A* exercised his or her volition (free will in a limited sense) to perform action *a*, but in effect action *a* is impossible without the preliminary conditions *C* (causal determinism). Either way, regardless of whether one is a soft determinist who makes minor accommodations for free will or a radical determinist who defines free will out of existence, an adherence to determinism entails an assent to some sort of mechanistic view of the universe and human action. From a rhetorical viewpoint, this preliminary adherence is translated into a larger framework or paradigm that the scientific rhetor imposes on his or her objects of investigation. The frame becomes anthropomorphized, despite its ostensible claims of objectivity, in that it is a human lens to understand the world and the lens in turn intersects with the realm of perception. Such a gesture is anthropomorphically mediated in the labyrinth of language, where rhetoric can be seen as a modest searchlight.

Dimensions of Anthropomorphizing

Martin Heidegger regards the curious relationship between technology and humans as follows:

> We ask the question concerning technology when we ask what it is. Everyone knows the two statements that answer our question. One says: Technology is a means to an end. The other says: Technology is a human activity. The two definitions of human technology belong together. For to posit ends and procure and utilize the means to them is a human activity. The manufacture and utilization of equipment, tools, and machines, the manufactured and used things themselves, and the needs and the ends that they serve, all belong to what technology is. The whole complex of these contrivances is technology. Technology itself is a contrivance...an *instrumentum*...the current conception of

> technology, according to which it is a means and a human activity, can therefore be called the instrumental and anthropological definition of technology.[20]

Technology remains a human endeavor. Yet one cannot speak meaningfully of anthropomorphizing without paying attention to its corollary – thinking of humans as machines. Although technology is a product of human endeavor, technology has the potential to define what the range of human endeavors is, or what it needs to be. Although - such a statement would imply some level of technological determinism – technocratic conceptions of society, for better and for worse, have come to dominate our conception of society and who we are. Technological determinism is rhetorically engineered with the 'computer' as the defining *leit motif* to characterize the *zeitgeist* of the age:

> As a calculating engine, a machine that controls machines, the computer does occupy a special place in our cultural landscape. It is the technology that more than any defines our age. Our generation perfected the computer, and we are intrigued by possibilities. Ruthlessly practical and efficient, the computer remains something fantastic...For us today, the computer constantly threatens to break out of the tiny corner of human affairs (scientific measurement and business accounting) that it was built to occupy, to contribute instead to a general redefinition of certain basic relationships: the relationship of science to technology, of knowledge to technical power, and, in the broadest sense, of mankind to the world of nature.[21]

The 'computer' is projected as the master metaphor that defines the cultural consciousness of the age. The computer becomes a lens with which to look at human nature. Sherry Turkle explains the inherent anthropomorphizing and subsequent 'thinking of yourself as a machine' phenomena emerging from the discourse of AI in a cogent manner:

> Computer models are seductive because they too put us in contact with issues that are both threatening and fascinating...People are afraid to think of

> themselves as machines, that they are controlled, predictable, determined...Similarly, although fearful, people want to find a way think about what they experience as the machine aspect of their natures; this is at the heart of the computer's holding power. Thinking about the self as a machine includes the feeling of being "run" from the outside, out of control because in the control of something beyond the self.[22]

As stated earlier, it is difficult to dissociate anthropomorphizing from 'thinking of yourself as a machine.' Even a redefinition of human nature in mechanistic terms, quite ironically, reflects on the ability of human desire to think of themselves as machines. From that perspective, even thinking of oneself as a machine is anthropomorphic. In Polanyi's words, such a gesture can be traced to the development of mathematical logic, to the extent to which, logical operations were considered co-extensive with the language of thought:

> The operations of the digital computers as *machines of logical inference* coincide with the operations of symbolic logic. We may therefore identify the formalization involved in the construction and the use of machines, operating in this particular way, with the procedure governing the construction of a deductive system. This procedure is threefold. (1) It designates undefined terms; (2) it specifies unproven asserted formulae (axioms); and (3) it prescribes the handling of such formulae writing down new asserted formulae (proofs). This result is achieved by a sustained effort to eliminate what are called 'psychological' elements – the factors which I call 'tacit.' The undefined terms are intended to stand without signifying anything, complete in themselves as marks on paper; unproven asserted formulae are to replace statements believed to be self-evident; operations constituting 'formal proof' are similarly intended to replace 'merely psychological' proof.[23]

The formalization of proof although a mathematical process in itself, has its philosophical and unformalized components as well, in that the idea of a proof entails an understanding of what a proof is and how it has to be executed and these 'unformalized supplements' are in the person's head:

> At a number of points, a formal system of symbols and operations can be said to function as a deductive system only by virtue of unformalized supplements, to which the operator of the system accedes: symbols must be identifiable and their meaning known, axioms must be understood to assert something, proofs must be acknowledged to demonstrate something, and this identifying, knowing, understanding, acknowledging are unformalized operations on which the workings of a formal system depends. We may call them the semantic functions of the formal system. These are performed by a person with the aid of the formal system, when the person relies on its use.[24]

The conceptualization of the formal system at some level should have room for including informal features as well: "the legitimate purpose of formalization lies in the reduction of the tacit coefficient to more limited and obvious informal operations; but it is nonsensical to aim at the total elimination of our personal participation."[25]

The next section will explore a couple of facets of anthropomorphizing namely a) attribution and b) identification within the discourse of mind design, which will help us better understand how this plays out in the discourse.

One of the salient features of anthropomorphizing in mind design literature is the principle of 'attribution,' achieved by establishing an isomorphism between humans and machines through the unifying lens of materialism. The idea of 'attribution,' hinges on a style of representation. Representation hinges on the principle of equivalence. Yet, representation originates in the human subject and emerges from the linguistic resources as a second-order depiction, which in turn is anthropomorphic at some level or the other. In *Representing and Intervening,* Ian Hacking makes a significant distinction between the world as it is, and the world as it is represented.[26] Unlike both Nelson Goodman or Richard Rorty[27] who advocate a radical anti-representationalism, Hacking reinstates representation, albeit, through an anthropomorphized lens:

> It will be protested that reality, or the world, was there before any representation or human language. Of course. But conceptualizing it as reality is secondary. First there is this human thing, the making of representations. Then there was the judging of representations as real or unreal, true or false, faithful or unfaithful. Finally comes the world, not first, but second, third or fourth...In saying that reality is parasitic upon representation, I do not join forces with those who, like Nelson Goodman or Richard Rorty, exclaim, 'the world well lost!' The world has an excellent place, even if not a first one. It was found by conceptualizing the *real as an attribute of representations* (emphasis mine)...I say that representing is curiously human. Call it species specific.[28]

Ascription or attribution is impossible without developing some mode of representation, whereby qualities belonging to a particular entity are projected onto another. By representation, one is referring to a depiction of likeness that is conferred on the object of investigation:

> Representations are first of all likenesses. Saying so flies in the face of philosophical truisms. There is, we all know, no representation without style...So it may be argued that there cannot be in the beginning have been simply representation, a creation of likeness. There must have been a style of representing before there was representing.[29]

The depiction of likeness is a rhetorical style of representing, before any representation can take place.

Turing's style of attribution is through the Turing test, whereby computers are attributed with intelligence if they can display intelligent behavior. Turing's imitation game is a representation of open-ended, task domains that requires normal interaction whereby the communicative behavior of the computer is not markedly dissimilar from that of the human. In fact, Turing considers one of the chief dissensions, namely the argument from consciousness to be largely solipsistic and irrelevant to his accounts of intelligence:

> According to the most extreme form of this view (that thought is impossible without consciousness) the only way by which one could be sure a machine thinks is to *be* the machine and to feel oneself thinking. One could then describe these feelings to the world, but of course no one would be justified in taking any notice. Likewise according to this view the only way to know that a *man* thinks is to be that particular man. It is in fact the solipsist point of view. It may be the most logical view to hold but it makes communication of ideas difficult. A is liable to believe 'A thinks but B does not' whilst B believes 'B thinks but A does not.' Instead of arguing continually over this point it is usual to have *the polite convention that everyone thinks*.[30]

Ascription of intelligence is largely based on verbal performance and a concurring isomorphism. Intelligence is defined as mental activity with a material basis. The anthropomorphizing inherent in Turing test should not be dismissed as a mere whim, since it can be located within a robust philosophical tradition of materialism (quite distinct from Cartesian substance dualism) that precedes Turing and the mechanistic praxis of Artificial Intelligence.

If intelligence is a completely distinct mental process that does not have a material basis, how can one build intelligent systems? Therefore, such an equation is necessary from the perspective of analogical reasoning. Once such an equation is established, AI researchers anthropomorphize machines using human like terms to describe mechanistic processes. Intelligence is ascribed to machines, as if it were internal to the machine in that the machine is ascribed with discrete mental states that are isomorphic with discrete mental states in the human computer at some level of abstraction.

Newell and Simon's physical symbol hypothesis rests on ascribing to a symbol system the ability to think and perform behavior that requires intelligence. Simon issues a declarative that computer simulations of thinking does indeed think: "a computer simulation of thinking thinks. It takes problems as it inputs and (sometimes) produces solutions as its outputs. It represents these problems and solutions as symbolic structures...and performs transformations on them like those the human mind does."[31] Simon's title for his essay "Machine as Mind," in itself is rhetorically insightful and provides a terministic screen with which to perceive the constituents of both human and machine intelligence: "the materials of thought are symbols-patterns, which can be replicated in a great variety of materials (including neurons and chips), thereby enabling physical symbol systems fashioned of these materials to think."[32] The underlying impulse behind Simon's model is that thought is essentially computation. Every thought process in the human mind has a corresponding neurological action, and by the same token, in a computer simulation "every thought process posited by a certain psychological theory has a corresponding process specified in the program," in such a decisive fashion that "computer simulations are thus directly parasitic upon some prior articulated theory about human psychology."[33]

Daniel Dennett's 'intentional systems' also articulates an ascription of anthropomorphic agency to machines: "intentional systems are obviously not all persons. We ascribe beliefs and desires to dogs and fish and thereby predict their behavior and we can even use the procedure to predict the behavior of machines."[34] Dennett's central thesis is that one can ascribe beliefs and desires to a computer in the manner in which one does for animals because the computer is an intentional system "not because it really and truly has beliefs and desires (whatever that would be), but just because it succumbs to a certain *stance* adopted towards it...the stance that proceeds by ascribing intentional predicates under the usual constraints of the computer, the stance that proceeds by considering the computer as a rational practical reasoner."[35] For instance, if a human were to begin a game with a queen's pawn opening, one can anticipate the chess-playing computer to rationally pick d7 to d5 or Nf6 among other openings that would give black (the chess playing computer) a reasonable control over the center, which is essential to any game of chess.

Either with Simon's physical symbol hypothesis that is a stronger version of anthropomorphism (whereby human thought is synonymous with computation) or with Dennett's intentional stance which is a milder version of anthropomorphizing (milder in a sense that the computer is deemed rational, not necessarily in an ontological sense but in a pragmatic sense for the sake of efficiency) the unifying thread is the principle of ascription. Ascription works as a first-order representation with Simon's model where human thought processes are synonymous with mechanistic computations or as a second-order representation with Dennett's model, where the prototype and the simulacrum are further removed.

The connectionists engage in a dialogic anthropomorphizing/machine-morphizing whereby the brain is considered equivalent to a machine and vice-versa. While their predecessors, the Symbolic AI researchers focused heavily on serial computing, the connectionists are driven by the 'neural net' metaphor, to focus on neural nets that owe their inspiration to the corresponding network of neurons in the human brain. The alliance of neurosciences with Artificial Intelligence founded on the framework of philosophical materialism gives birth to a taxonomy of new concepts and theories that are used interchangeably while speaking about human intelligence and natural intelligence. Even the very nomenclature that is used, from a rhetorical perspective, orients the audience towards a particular set of goals, namely the understanding of the human 'self' in terms of the 'machine' and an understanding of the 'machine' in terms of the 'self.' Churchland argues for the viability and dominance of neuroscientific/computational accounts of mind as the framework with which to understand our inner nature:

> If materialism, in the end, is true, then it is the conceptual framework of a completed neuroscience that will embody the essential wisdom of our inner nature…Consider then the possibility of learning to describe, conceive and introspectively apprehend the teeming intricacies of one's inner life within the conceptual framework of a 'completed' neuroscience, or one advanced far beyond its current state. Suppose we train our native mechanisms to make a new and more detailed set of discriminations, a set that corresponds not to primitive psychological taxonomy of ordinary language, but to some more penetrating taxonomy of states drawn from a 'completed' neuroscience. And suppose we train ourselves to respond to that reconfigured activity with judgments that were framed, as a matter of habit, in the appropriate concepts from neuroscience…I suggest then, that the genuine arrival of a materialist kinematics and dynamics for psychological states and cognitive processes will constitute not a gloom in our inner life is eclipsed or suppressed, but rather a dawning, in which its marvelous intricacies are finally revealed—even, if we apply ourselves, in self-conscious introspection.[36]

To this, this writer might add – 'neurons of the world unite, all you have to lose is your obscurity, and the shackles imposed by folk psychology.' In fact, the interest generated in connectionist models both in academia and the industry has been phenomenal, to say the least. Instead of relying on serial computing, connectionist, neuroscientific models are based on the concept of neural nets. The human brain in turn is a neural net, an aggregate of interconnected nerve cells, that in turn constitutes the substrate for intelligent behavior and also embodies what folk psychologists depict as the 'mind.'

The biological 'neural net' became the overarching metaphor or anthropomorphic lens to ascribe 'intelligence' to the mechanistic 'neural net.' In the words of Bolter:

> Today the computer is constantly serving as a metaphor for the human mind or brain: psychologists speak of the input and output, sometimes even the hardware and software, of the brain; linguists treat human language as if it were a programming code; and everyone speaks of making computers think...A defining technology develops links, metaphorical or otherwise, with a culture's science, philosophy, or literature; it is always available to serve as a metaphor, example, model, or symbol. A defining technology resembles a magnifying glass, which collects and focuses seemingly disparate ideas in a culture into one bright, sometimes piercing ray. Technology does not call forth major cultural changes by itself, but it does bring ideas into a focus by explaining or exemplifying them in new ways to larger audiences.[37]

Warren S. McCulloch and Walter S. Pitts theoretically elaborated on the feasibility of representing activity in neural nets in terms of symbolic propositions.[38] McCulloch and Pitts' 'neural nets' are equivalent to universal Turing machines, in that all mental processes could be described by a finite number of symbolic expressions "embodied in nets of what they called "formal" neurons."[39] The notion of neural nets gave birth to Parallel Distributed Processing. Instead of having a single processing unit, not unlike the network of interconnected nerves in the brain, parallel processing relies on a network of a number of processors operating simultaneously. Several doubts persist (need citation here) whether artificial neural nets do indeed simulate biological nervous systems, the conception of a neural net in computational terms is closer to realistic modeling of biological systems than serial computing.

The striking feature of connectionism is the reversal of metaphors compared to the scheme employed by symbolic AI, namely connectionism represents the brain as a computer, while symbolic AI represents the computer as a brain. The computational account of mind under serial or symbolic computing did not shed enough insight into the actual processing of mental activity in the brain. Symbolic AI is a superimposition of a computational model on the human mind. The connectionists, on the other hand, claim to enhance the range of tasks of machine intelligence as well as shed more insight into the human brain itself. David E. Rumelhart claims:

> The basic strategy of the connectionist approach is to take as its fundamental processing unit something close to an abstract neuron. We imagine that computation is carried out through simple interactions among processing units. Essentially the idea is that these processing elements communicate by sending numbers along the lines that connect the processing elements...*The operations in our models then can be characterized as "neurally-inspired"* (emphasis mine). How does the replacement of the computer metaphor with a brain metaphor as model of mind affect our thinking? This change in orientation leads us to a number of considerations that further inform and constrain our model-building efforts. Perhaps the most crucial of these is time. Neurons operate in the time scale of milliseconds, whereas computer components operate in the time scale of nanoseconds...this means that human processes that take on the order of a second or less can involve only a hundred or so time steps...The use of brain-style computational systems, then, offers not only the hope that we can characterize how brains actually carry out certain information-processing tasks but also solutions to computational problems that seem difficult to solve in more traditional computational frameworks. It is here where the ultimate value of connectionist systems must be evaluated.[40]

Rumelhart does at least two things: a) argues for equivalence between neurons and the processing units and b) asserts that neurons are slower and hence involves fewer steps when compared to the processing abilities of the computer. The implicit assumption here is that neurons act chiefly electrically, while other chemicals may come into play as well. Although from a mathematical perspective Rumelhart's argument makes good sense when one takes into account the efficacy of high-speed processing units, the equation of the nature of neural activity in the brain and parallel-processing is not that explicit. Furthermore, the greater speeds of processing units by themselves do not guarantee that they can accomplish the repertoire of mental activities that slower neurons are engaging in. Yet, from a heuristic standpoint, the anthropomorphic projection of the biological brain as a model for computation is useful in that for a set of neatly defined domains parallel processing models are generally considered more efficacious than serial computing models. (PDP and other pertinent concepts of connectionism will be explained in greater detail in chapter 6).

Although the means of symbolic AI and connectionist AI are different, from a rhetorical point of view, there is a unifying thread that connects the two in that both provide us with computational models of mind. Like 'iron sharpens iron,' computational models shed light on human cognition and vice-versa. Both models are representations of mental activity, considered isomorphic in structure and anthropomorphic in terms of rhetorical appeal. The models by themselves, be it the brain as a computer or the computer as a brain, are superimpositions of anthropomorphized models in the ordering of reality, to such an extent that the ensuing details are rhetorically delineated to respectively fit the story that each paradigm seems to be telling. The models are also mirrors in that one can indeed empirically find evidence for demonstrating such equivalence, with respect to certain domains. The models become styles of representation.

Each period in human history has come up with its own styles of representations. Erich Auerbach's view that mimesis takes on varying, context-dependent, representative styles in western literature is salient for the history of ideas as well.[41] How we represent is influenced by the dominant intellectual currents of our times? Would connectionist AI have enjoyed the same respectability were it not for the rapid ascendancy of neuroscientific models of mind? By the same token, one can only speculate about what types of models are going to dominate conceptions of mind a few generations from now.

Byron Reeves and Clifford Nass suggest that three dominant views have emerged when people talk about human-computer relationships: a) one-down, b) one-up and c) one across.[42] The one-down position argues that computers are mere tools and should be one down with regard to the human user. A weaker AI thesis would probably articulate this position. The one-up position posits the computer as "master:" "the computer should take charge and absorb as much of the work is possible...under this view, computers are "wizards," "autonomous agents," or "guides."[43] Computational models of mind would probably fall under the computer as "master" category, since there is an essential redefinition of human nature when viewed through the lens of computation, as contrasted with traditional humanistic models. The one-across position emphasizes cooperative partnership that acknowledges that humans and machines are mutually dependent on each other. Most of the literature surveyed in mind design, especially those that are not antithetical to the development of Artificial Intelligence would either fall under the 'one-up' or 'one-across' camps. Much of pro-AI literature advocates robust pictures of machine intelligence, where machines are endowed or attributed with a personality that enhances the rhetorical appeal of machine intelligence.

Reeves and Nass articulate that in the arena of human-computer interaction:

> *Perceptions* are far more influential than reality defined more objectively. When perceptions are considered, it doesn't matter whether a computer can really have a personality or not. People perceive that *it* can, and they'll respond socially on the basis of perception alone...these responses are unfairly labeled as irrational. They are merely human, and part of any communication experience.[44]

The perception of personality that is ascribed to machines arises from a closely related concept, namely, identification. In some ways, the perception of personality is also aesthetic in a sense that it deals with the subjective aspects of human experience. Subjective frames are superimposed on machines so that humans can achieve a sense of familial kinship during the course of the interactions. Barry Brummett avers that the "subjective aesthetic experience is one of identification, of investment of the self with the object or action; it become us, and we become part of it," and is "particularly appropriate to machine aesthetics...It is the appreciation of a table saw as it does cut wood, producing smooth planes and precise angles again and again in perfect order."[45] The machine, *qua* machine is not merely an instrumental device whose value is measured by its ability to deliver certain outputs at a particular moment, but has indefinitely become the mainstay of (post)modernity in that its position is not just one of an object, but also as a culturally, evocative subject. An evocative subject spells out a teleological orientation, an orientation that necessitates an understanding of its indispensability but also an accompanying missive as to how humans should interact with it. The computer metaphor has pervaded the cultural consciousness of our time, engulfing academia, industry and popular culture as a whole. The increasing computerization of various segments of society involves a human investment in the technology, to an extent, that we are beginning to see technology as extensions of ourselves attempting to finish the task that we have begun.

With rapidly accelerating levels of sophistication, computers are increasingly being perceived as autonomous agents whose performance is relying less and less on human interventions. Even so, as Kenneth Burke notes with regard to "autonomous" activities, the principle of "Rhetorical identification" may be summed up thus: The fact that an activity is capable of reduction to intrinsic, autonomous principles does not argue that it is free from identification with other orders of motivation extrinsic to it."[46] Rhetorical identification emerges as a symbolic action, whereby the human agent identifies with the autonomous agent in its ability to demonstrate intelligent behavior.

Identification revolves around at least two principles, namely, salience and resonance. The Turing test is an anthropomorphic identification, in a sense, that conversational abilities are salient to humans, and resonates with our deepest needs to make contact. Speech and writing have become central to our self-definition, for it is in our ability to communicate that we see who we are, our place in this world, and our relative positioning with other individuals and objects. And 'no man is an island' indeed. If Artificial Intelligence evolves to live up to Turing's promises and prognostications in passing the Turing test, one could chat with computers on an equal footing, that is, one agent talking to another. The range of topics may run from mathematics to literature. Under these circumstances, the corresponding anthropomorphized identification is only going to rise significantly. It might be expedient to interject that anthropomorphizing are not due to a mistaken human gullibility, but due to our proclivity to symbolize and envision the world through lenses that are most amicable to 'sense-making.'

The rhetorical identification is not necessarily of the type A = B, instead it only presupposes that A and B share a certain symmetry with each other, that makes a qualitative mapping of attributes into each other possible. The identification via an isomorphism is a plausible position even if it dampens the celebratory chorus that humans are entirely equivalent to machines. One could say that humans are machines of some sort, but yet the sense in which the human is a machine is not quite the same as the assertion that automatons are machines, due to differences in architecture and bio-chemical make up and other metaphysical issues as well which will not be addressed in this discussion.

Despite the difference, the process of identification becomes salient when one specifies the simulatable domains of human activity that can be translated, programmed, and implemented by machines. Interestingly enough, once these tasks are accomplished, the way in which humans perform these tasks are interpreted through the lens of a machine. In other words, the purveyors of our intellectual culture have not escaped the panopticon of the technological gaze, and have in turn donned the priestly trappings of the technological church in creating a significant role-reversal. Namely, the creative subject (the human) while interacting with technology (the artifact) has been recast as the object (the human), redefined by computational models (the creative subjects) of human nature. Regardless of whatever thesis one subscribes to, be it the stronger or weaker AI thesis, the salience and centrality of computer technology in modern life has created an ineluctable sense of identification with technology, that only a Luddite-style retreat into the pristine backwoods of the pre-scientific human can avert the imperialistic intrusion of computers into our everyday life. However, such a retreat is inconceivable and unlikely since even the so-called 'developing world' (usually romanticized in the West as being less materialistic and less preoccupied with technology) has for better and for worse, jumped on the bandwagon of the digital revolution.

The suasory appeal of intelligent machines with its almost flawless rationality and efficiency creates a strong sense of identification with the 'scientific' telos of what a human ought to *be*. Especially, the early part of the digital revolution was concerned primarily with the domains of logic, mathematics and related fields. Recently, there has been a growing interest in emotive/affective computing as well (this topic will be covered in a study at a later period of time). In other words, even emotive or subjective qualia are beginning to be scrutinized through computational lenses. To summarize: Digital determinists are earnest about making humans fully 'computerized.' In the same breath, the obverse reflection depicts an equally compelling desire to make computers humanized.

The case for identification here is not necessarily uniform across all types of human-machine interactions, but more pronounced with machines that are 'smarter.' Identification primarily works between objects where we find sufficient likenesses or family resemblances. Ludwig Wittgenstein enunciates that definitions assigned to objects are largely based on identification and family resemblances: "I can think of no better explanation to characterize these similarities than "family resemblances;" for the various resemblances between members of a family: build, features, color of the eyes, gait, temperament, etc. etc. overlap and criss-cross the same way."[47] He goes on to add:

> Why do we call something a "number"? Well, perhaps because it has a – direct – relationship with several things that have hitherto been called numbers; and this can be said to give it an indirect relationship to other things we call the same name. And we extend our concept of number as in spinning a thread we twist fiber on fiber. And the strength of the thread does not reside in the fact that some one fiber runs through its whole length, but in overlapping of many fibers.[48]

To apply Wittgenstein's insights, one can say the nomenclature 'Artificial Intelligence,' especially with regard to the term 'intelligence' is a family resemblance of identification in that one sees a range of tasks performed by the human prototype that can be simulated in machine models. The Turing test, the physical symbol hypothesis, the intentional stance, artificial neural nets among other nomenclatures in AI literature are identifications assigned on family resemblances.

Are these family resemblances based on ontological realism? That, this chapter is ill equipped to address. However, there are teleological similarities in the discourse of mind design, namely, in the pursuit of common theoretical and practical goals – cast as the understanding of the human through 'machine eyes' and the understanding of the machine through 'human eyes.' Indeed, the mechanization of the human (sometimes perceived as an effacement) is in itself the fulfillment of an anthropomorphized desire, namely, the desire to find facts to fit the hypotheses of a particular mechanistic, philosophical orientation. Irony of ironies – in that a move to completely 'mechanize' humans, is on further reflection, a human desire in itself.

Joseph Weizenbaum suggests that the metaphorical elements and the connotations associated with the term "machine," makes identification with machines more feasible – the predictive regularity of machines make machines more reliable when it comes to discussing the 'mind' (a product of the biological brain) and the transforming power of the computer:

> The stretching of connotative range of the word "machine" has two quite separable significances: First, it testifies that folk wisdom recognizes the essential characteristic of the machine to be its relentless regularity, its blind obedience to the law of which it is an embodiment...This is the insight which permits people to talk of, say, a bureaucracy or a system of justice as a machine. Second, it reveals an implicit, though very vague, understanding in the folk wisdom of the idea that one aspect has to do with information transfer and not with the transmission of material power. The arrival of all sorts of electronic machines, especially of the electronic computer, has changed our image of the machine...to that of a tranformer of *information*.[49]

The calculus of bounded and flawless rationality and the machine's ability to process voluminous amounts of information resonates with the scientific image of the human. The emphasis on scientific rationality has made identification more amenable, especially in circles where technical rationality is an undisputed virtue.

Certainly, there is indeed tangible basis for asserting that mimetic identifications or simulations of intelligent behavior are possible once we specify the task domains and make sure that the 'simulatable' tasks are operationalized. Such an enterprise, at surface level, may even appear 'objective' and 'empirical.' Yet, one's value orientation comes into play as well. In other words, agent-relative orientations are foregrounded in the style and make-up of the artifact. The values that are most salient and resonate with the individual are most reflected in the artifact. Individuals who are exceptionally proficient in mathematics or any of the sciences are most likely to impose scientific frames on the world around them, since a scientific orientation is more salient to them. By the same token, artistic individuals are likely to impose aesthetic frames on the world around them, since an aesthetic orientation is more salient to them.

Concluding Remarks

The inescapable presence of human motives, makes it difficult if not impossible, to think outside of ourselves. As long as human motives exist, anthropomorphizing is here to say. On a precautionary note, it might be helpful not to confuse anthropomorphisms with anthropocentrism since the latter view borders on arrogance.

Anthropomorphizing is inherent is our models which are at best incomplete descriptions of reality. However, the looming specter of incompleteness should not discourage us from exploring, as long as we are willing to recognize our limitations. Incompleteness only attests to the condition of us being finite in this world. In closing, the words of the pre-Socratic philosopher Empedocles best illustrates the sentiment this chapter has been trying to express:

> For limited are the means of grasping *(i.e. the organs of sense-perception)* which are scattered throughout their limbs, and many are the miseries that press in and blunt the thoughts. And having looked at (*only*) a small part of existence during their lives, doomed to perish swiftly like smoke they are carried aloft and wafted away, believing only that upon which as individuals they chance to hit as they wander in all directions; but every man preens himself on having found the Whole: so little are these things to be seen by men or to be heard, or to be comprehended by the mind.[50]

Notes

[1] Michael Polanyi, *Personal Knowledge: Towards a Post-Critical Philosophy* (Chicago: The University of Chicago Press, 1974), 3.
[2] By no means is this a radical or new formulation. However, this view still receives significant resistance especially in circles that are committed to the view of a disengaged or impersonal knowledge. Removing the 'person' from the discussion is after all the most reliable means of achieving knowledge – or so, we are told.
[3] Polanyi, *Personal Knowledge*, 60-61.
[4] Polanyi, *Personal Knowledge*, 163.
[5] St. Augustine, *The Confessions of Saint Augustine Book XI*, trans. Rex Warner (New York: Signet, 2001), 274-276.
[6] Augustine, *Confessions*, 276.
[7] Georges Lemaitres, "The beginning of the world from the point of view of quantum theory," *Nature* 127 (1931): 706.
[8] Helge Kragh, *Cosmology and Controversy: The Historical Development of Two Theories of the Universe* (Princeton: Princeton University Press, 1996), 80.
[9] G. Gamov and C. L. Critchfield, *Theory of Atomic Nucleus and Nuclear Energy-Source* (Oxford: Clarendon Press, 1949).
[10] Fred Hoyle, *The Nature of the Universe* (Oxford: Blackwell, 1950).
[11] F. Hoyle and R. A. Lyttleton, "A New Model for the Expanding Universe," *Monthly Notices of the Royal Astronomical Society* 108 (1948): 372-382.
[12] Hoyle, *The Nature of the Universe*, 124.
[13] H. A. Bethe, "Energy Production in the Stars," *Physical Review* 55(1939): 434-456.
[14] William A. Fowler, "The Quest for the Origins of the Elements," *Science* 226 (1984): 922-935.
[15] Participant observation may also be seen as a form of scientific ethnography. However, I refrained from using the term 'ethnography' because I have not yet thought through the implications of using the term 'scientific ethnography.'
[16] Polanyi, *Personal Knowledge*, 15-16.
[17] cited in Gerald Holton, *The Scientific Imagination* (Cambridge, Massachusetts: Harvard University Press, 1998), 203.
[18] Ludwig von Bertalanffy, *Perspectives on General Systems Theory*, ed. Edgar Taschidjian (New York: George Braziller, 1975), 49.
[19] Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago: The University of Chicago Press, 1996).
[20] Martin Heidegger, *The Question Concerning Technology and Other Essays*, trans. William Lovitt (New York: Harper & Row, 1977), 4-5.
[21] J. David Bolter, *Turing's Man: Western Culture in the Computer Age* (Chapel Hill: The University of North Carolina Press, 1984), 8-9.
[22] Sherry Turkle, *The Second Self: Computers and the Human Spirit* (New York: Simon and Schuster, 1984), 299.
[23] Polanyi, *Personal Knowledge*, 258.

[24] Polanyi, *Personal Knowledge*, 258.
[25] Polanyi, *Personal Knowledge*, 259.
[26] Ian Hacking, *Representing and Intervening: Introductory topics in the Philosophy of Natural Science* (Cambridge: Cambridge University Press, 1983).
[27] Nelson Goodman, *Ways of Worldmaking* (Indianapolis: Hackett Pub, 1978); Richard Rorty, *Philosophy and the Mirror of Nature* (Princeton: Princeton University Press, 1979).
[28] Hacking, *Representing and Intervening*, 136.
[29] Hacking, *Representing and Intervening*, 137.
[30] A. M. Turing, "Computing Machinery and Intelligence," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge: Bradford, MIT Press), 29-56.
[31] Herbert Simon, "Machine as Mind," in *The Legacy of Alan Turing, vol.1*, ed. Peter Millican and Andy Clark (Oxford: Oxford University Press, 1999), 81-102 (cited in page 82)
[32] Simon, "Machine as Mind," 82.
[33] Margaret A. Boden, *Artificial Intelligence and Natural Man* (Sussex: The Harvester Press, 1977), 5.
[34] Daniel Dennett, *Brainstorms* (Brighton: Harvester Press, 1981), 271.
[35] Dennett, *Brainstorms*, 271-272.
[36] Paul Churchland, *Matter and Consciousness* (Cambridge: A Bradford Book, MIT Press, 1999), 179-180.
[37] Bolter, *Turing's Man*, 11
[38] Warren S. McCulloch and Walter S. Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," in *The Philosophy of Artificial Intelligence*, ed. Margaret A. Boden (Oxford: Oxford University Press, 1990), 22-39.
[39] Jack D. Cowan and David H. Sharp, "Neural Nets and Artificial Intelligence," in *The Artificial Intelligence Debate: False Start, Real Foundations*, ed. Stephen A. Graubard (Cambridge: MIT Press, 1988), 84-121.
[40] David. E. Rumelhart, "The Architecture of Mind: A Connectionist Approach," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge: Bradford, MIT Press), 205-232.
[41] Erich Auerbach, *Mimesis: The Representation of Reality in Western Literature*, trans. Willard R. Trask (Princeton: Princeton University Press, 1953).
[42] Byron Reeves and Clifford Nass, *The Media Equation: How people treat computers, television and new media like real people and places* (Cambridge: Cambridge University Press, 1996).
[43] Reeves and Nass, *The Media Equation*, 159.
[44] Reeves and Nass, *The Media Equation*, 253-254.
[45] Barry Brummett, *Rhetoric of Machine Aesthetics* (Westport, Connecticut: Praeger, 1999), 19-20.
[46] Kenneth Burke, *A Rhetoric of Motives* (Berkeley: The University of California Press, 1969), 27.
[47] Ludwig Wittgenstein, *Philosophical Investigations*, trans. G.E.M.Anscombe (New York: Macmillan, 1968), 32.
[48] Wittgenstein, *Philosophical Investigations*, 32.
[49] Joseph Weizenbaum, *Computer Power and Human Reason: From Judgment to Calculation* (New York: W.H.Freeman and Company, 1976), 41.
[50] Empedocles, 'Fragments,' in Kathleen Freeman, *Ancilla to the Pre-Socratic Philosophers* (Cambridge, Massachusetts: Harvard University Press, 1996), 51.

CHAPTER IV

IMAGO MACHINA: ALAN TURING AND RHETORIC

> "Man is heaven-born; not the thrall of Circumstances, of Necessity, but the victorious subduer thereof" – Boswell's *Life of Johnson.*[1]

> "All the great…things that have been achieved in the world, have been achieved by individuals" – Coleridge, *Table Talk*[2]

> The great Fact of Existence is great to him. Fly as he will, he cannot get out of the awful presence of this Reality. His mind is so made, he is great by that, first of all. Fearful and wonderful, real as Life, real as Death, is this universe to him (emphasis mine). Though all men should forget its truth, and walk in a vain show, he cannot. At all moments the Flame-image glares in upon him; undeniable, there, there! - I wish you to take this as my primary definition of a Great Man. A little man may have this…but a Great Man cannot be without it" – Thomas Carlyle, On Heroes, Hero-Worship, & the Heroic in History.[3]

Thomas Carlyle's "Great Man" theory of history although poetically inspirational is perhaps found philosophically wanting when applied to the transmission and germination of the history of ideas. "No man is an island" indeed. The overpowering impulse to accord one individual non-pareil originality although quite heroic tends to neglect the institutional or social conditions that helped foster the birth of innovative ideas. It is therefore safe to assume that hero-worship is an insidious and jejune sentiment. Insidious because it neglects the collectivity, and jejune because it is a callow assessment of how change is conceived and disseminated in human societies. Yet, the human imagination is impoverished and cannot attain creative entelechy without the spirited narratives of the 'heroic' spurring us onward. The image of the 'heroic' is an alluring and iconic one. In the absence of the heroic, what is left is the proclivity to comfortlessly 'fall into quotidian.' The ordinariness of mere existence and obscurity is less rhetorically compelling in stark contrast to the narratives of extraordinary fortitude or brilliance, characteristic of renowned thinkers who - as Thomas Kuhn would describe -created a paradigm shift.[4] Copernicus, Newton, Einstein, to name a few, redefined the way in which we perceive the colossal universe around us. However, without recklessly capitulating to a 'great person' account of intellectual history or by the same token, obliterating the individual narratives of excellence and originality, it is useful to give credit where credit is due.

The field of Artificial Intelligence has its heroes. To restate the obvious, computers have engulfed just about every aspect of human organization. Any attempt to articulate a comprehensive, conceptual account of the emergence of computing is deeply intermeshed with the voluminous history of mathematics itself. A teeming ocean with multitudinous faces appear and recede; first in prominence and then in obscurity as successors improvise and innovate beyond what was originally conceived. It is safe to conjecture that any anthology of the history of artificial intelligence is incomplete without reference to the enigmatic and heroic persona of Alan Turing.[5] Turing's pioneering work in Artificial Intelligence is a direct predecessor (if not the actual fountain-head) for the computer revolution in the twentieth century.

In spite of his life of relative obscurity, the remarks of Turing's biographer Andrew Hodges are especially pertinent, with regard to justifying the salience of his work:

> For mathematicians he has immortality through the expression of the Turing machine. Many people must have used his name without any conception of his historical existence – the nearest thing to life as a disembodied spirit that he

> once pondered on. Going even further, modern papers sometimes employ the usage of *turing machines*. Sinking without a capital letter into the collective mathematical consciousness (as with the *abelian group*, or the *riemannian manifold*) is probably the best that science can offer in the way of canonization.[6]

By conceiving the idea of a universal machine (composed of discrete states that could perform tasks that would normally require intelligence if carried out by the human operator) Turing essentially articulated a blueprint for machine intelligence. Turing can be seen as a precursor to both symbolic and connectionist AI, not to mention other significant advances in computing including quantum computing and programmable matter that might at some point at least acknowledge some level of indebtedness to Turing's legacy. In his germinal essay, "Computing Machine and Intelligence" Turing reflects on the question 'can machines think?"[7] To circumvent the natural tendency to respond to such a question with a simple yes or no, Turing exercises rhetorical astuteness by replacing the question "can machines think?" with "are there imaginable digital computers which *would do well* in the imitation game?"[8] Turing's strategy in replacing an ontological question with a more pragmatic and behavoristic question is rhetorically noteworthy. It replaces grappling the concept of 'thinking' as an ontological entity with the idea of acting like 'thinking.' Set against this backdrop, the purpose of this chapter is two-fold a) to trace the rhetorical situation that facilitated the development of Turing's blueprint for machine intelligence and b) undertake a Burkean analysis of Turing's strategy in arguing the case for machine intelligence using Burke's conception of a terministic screen and simultaneously discuss Turing's metaphor of mind. A close reading of Turing's texts help uncover

Turing's assumptions of the human mind and the subsequent impact these assumptions play in building intelligent agents– furthermore, Turing's rhetoric also sheds light on a tradition of thinking even larger than the AI community, namely the physicalist view of the human mind and human nature that is undoubtedly pervasive and dominant in both the sciences and the arts. Before undertaking a rhetorical analysis of Turing's underlying assumptions, tracing the intellectual contours of Turing's evolution might illuminate and contextualize our understanding of the philosophical origins of artificial intelligence.

Rhetorical Situation

Since the concept of a rhetorical situation is folklore in the annals of modern rhetorical theory, much elaboration is not needed. This writer will employ the term rhetorical situation synonymously with rhetorical and historical contexts that lend certain ideas more persuasive credence than others. It is possible to look at Turing's ideas in isolation, purely as an exegetical exercise in textual criticism. Such an approach might have hermeneutic depth in teasing out the complexities of the text; on the other hand, it would ignore the 'situatedness' of ideas, how ideas chain out either as part or as a reaction against the *zeitgeist* of a time. In that spirit, this writer locates Turing within a larger physicalist/behaviorist paradigm of mind, whose conceptions of mental phenomena in varying forms have come to dominate conceptions of mind in much of the twentieth century. The rhetorical situation underlying Turing's thought (in other words, what are the contexts that facilitated Turing's enunciation of his position) will be explicated through the following four sections entitled: a) mind-body and electronic brains, b) biochemical materialism and minds and c) logical and mathematical origins (this particular section tries to show how Turing's interest in computers was inspired by grappling with theoretical challenges in mathematics).

An attempt to trace the evolution of Turing's thought with regard to machine intelligence is a long and circuitous one, which may include biographical material that goes back to his formative years. The young Alan Turing even as an adolescent embraced the view that Edwin Tenney Brewster advocated in *Natural Wonders*: "For, of course, the body is a machine. It is a vastly complex machine, many, many times more complicated than any machine ever made with hands; but still after all a machine."[9] Brewster advanced the view that the human body was a biochemical machine and that human behavior could be explained naturalistically without resorting to the dualism between the mind and body implicit in Plato and more explicit in Descartes. On a more personal note, it is safe to hypothesize that one of the factors that played a huge role in Turing's intellectual development is his intimate and very moving, adolescent friendship with Christopher Morcom, a high school friend. Turing found in Christopher Morcom a very astute partner with whom he could discuss mathematics, physics among other topics that played an influential role in shaping Turing's intellectual development. Morcom's untimely death deeply impacted Turing both intellectually and theologically, leading him to ponder weightier issues pertaining to the relationship between the mind and the body, the spirit and the body. In some ethereal and non-tangible sense, Turing felt connected to the departed spirit of his bosom companion, leading him to wonder about the possibility of constructing a nonmaterial 'spirit' like a "wireless set resonating to a signal from the unseen world?"[10] In a letter to Mrs. Morcom, Turing grapples with the mind-body, spirit-body relationship and attempts to reconcile the apparent dichotomy in the following manner:

> Personally I think that spirit is really eternally connected with matter but certainly not always by the same kind of body. I did believe it possible for a spirit at death to go to a universe entirely separate from our own, but I now consider that matter and spirit are so connected that this would be a contradiction in terms. It is possible however but unlikely that such universes may exist. Then as regards the actual connection between spirit and body I consider that the body by reason of being a living body can 'attract' and hold on to a 'spirit', whilst the body is alive and awake the two are firmly connected. When the body is asleep I cannot guess what happens but when the body dies the 'mechanism' of the body, holding the spirit is gone and the spirit finds a new body sooner or later perhaps immediately. As regards the question of why we have bodies at all; why we do not or cannot live free as spirits and communicate as such, we probably could do so but there would be nothing whatever to do. The body provides something for the spirit to look after and use.[11]

Later on, Turing completely eschewed the 'spirit' and embraced a wholesale materialism, however his adolescent speculations in metaphysics and his desire to construct an 'electronic brain' as a tribute to Christopher Morcom[12] permanently captivated his interest in the quest for building intelligent machines.

The early Turing before his conversion to anti-supernatural materialism appears to have been influenced by Sir Arthur Eddington. Eddington rejected a strict Laplacian determinism that appeared to him as an unreflective application of Newton's Laws of motion. The Laplacian view asserted that:

> An intelligence that, at a given instant, could comprehend all the forces by which nature is animated and the respective situation of the beings that make it up, if moreover it were vast enough to submit these data to analysis, would encompass in the same formula the movements of the greatest bodies of the

> universe and those of the lightest atoms. For such an intelligence nothing would be uncertain, and the future, like the past, would be open to its eyes.[13]

If one were to know the positions and velocities of every particle in the universe, one could simply use Newton's laws of motion to determine at least theoretically their positions or velocities either in the past, the present or the future. (Even in theological circles, there was a variant of this argument in the debate between Calvin's theory of divine predestination and Armenius' account of free will – in secular philosophical circles, the argument is framed as the contest between free will and biological determinism). As Brian Greene explains, "this rigid lock-step view of the unfolding of the universe raises all sorts of perplexing philosophical dilemmas surrounding the question of free will, but its import was substantially diminished by the discovery of quantum mechanics."[14] Heisenberg's Uncertainty Principle and the now renowned Schrödinger's Cat thought-experiment among others undermined the Laplacian view by arguing that one could not assert with certainty the precise positions and velocities of the physical constituents in the universe. Instead of resorting to the classical properties of the Newtonian universe, the quantum physicists asserted the probabilistic nature of wave functions.

Eddington surmised that the open-ended nature of the wave function, opened up the doors for accomodating 'free will,' and consequently eclipsed the mechanistic clockwork view of the universe. In Eddington's words:

> I have an intuition much more immediate than any relating to the objects of the physical world; this tells me that nowhere in the world as yet is there any trace of a deciding factor as to whether I am going to lift my right hand or my left. It depends on the unfettered act of volition not yet made or foreshadowed.[15]

Implicit in Eddington's conception is the suggestion that the mind could control the brain, not necessarily in a simplistic sense such as the manipulation of a wave function that could lead to a mental decision. By rehabilitating the notion of free will, Eddington mused, "it seems that we must attribute to the mind power not only to decide the behavior of atoms individually but to affect systematically large groups – in fact to tamper with the odds on atomic behavior."[16]

Turing considered Eddington's ideas to bridge the gap between the mechanistic view of the body that he inherited from Brewster and the idea of the 'spirit' that he wanted to believe in. McTaggert's idealism, Eddington's free will theories and Brewster's mechanism of the body combined to form the background of Turing's assertion that "We have a will which is able to determine the action of the atoms probably in a small portion of the brain, or possibly all over it."[17] Although, there is no single, definable moment of epiphany, the adult Turing veered away from both idealistic theories and notions of free will and embraced biological determinism wholesale.

Biochemical Materialism and Minds

Thomas Hobbes in the *Leviathan* directly refutes Cartesian substance dualism by arguing explicitly that thinking is a mechanical process (not to mention materialist thinkers in France during and after Descartes' time). A materialist philosophy of mind laid the groundwork for conceiving the mind as a machine. However, there have been significant variations revolving around a common theme. Re-telling all the different versions of materialism in this section is not necessary. Only a few salient examples that make explicit connections between materialism and the mind will be surveyed here. As early as the late 1800s with the advent of Darwinian naturalism and scientific materialism, William Graham speaks of a science of mind in *The Creed of Science* as follows:

> Consciousness, is only...an accidental bye-product (*sic*) – something over and above the full and fair physical result, which by an accident, fortunate or otherwise, appeared to watch over and register the whole series of physical processes, though these would have bone (*sic*) on just as well in its absence.[18]

Another important example that comes to mind in the 1900s that bridges the brain sciences and mechanistic thinking, is the claim advanced by Jacques Loeb who surmised, "what the metaphysician calls consciousness are phenomena determined by the mechanisms of associative memory."[19] Borrowing the notion of *tropism* from Julian Sachs' work in botany, Loeb advanced the idea of animal heliotropism to demonstrate that certain organisms are merely photochemical machines responding to the external stimuli of light. Loeb demonstrated that when caterpillars of *Porthesia chrysorrhoea* are placed in front of a stream of light coming from a direction that is directly opposite to the supply of food, they invariably move towards the light and not the food and perish as a result. This experiment can be used as an illustration to undermine the metaphysical notion of the will, that all creatures are chiefly dictated by the instinct for self-preservation: "In this instance the light is the "will" of the animal which determines the direction of its movement, just as it is gravity in the case of a falling stone or the movement of a planet"[20]

Shanker elucidates Loeb's argument that "heliotropic animals" are "photometric machines" as a robust argument for eliminative materialism that demystifies some of the mentalist, teleological assumptions of the mind:

> The fact that caterpillars expired for want of food was no more a demonstration of (perverse!) purposive behavior than the converse result would have supported vitalism. To suppose that their motor responses could exhibit the complexity of human purposive behavior is once again to assume *ab initio* that *intentions* and *volitions* are simply part of a causal chain, from which the ability to choose, decide, select and deliberate are excluded a priori. But that was exactly what Loeb intended! This was not to be an isolated attack on the notion of *will*: all of the 'mentalist' concepts were to be removed from the eliminativist analysis of purposive – equals self-regulating – behavior.[21]

When purposive behavior, usually cast in metaphysical accounts of the "will" is recast in neurological terms, eliminative materialists suggest that one could do away with speculative metaphysics and thus resort to a purely neurological account of human cognition and behavior. If this were the case, even terms such as consciousness, creativity, will and so forth are construed purely in neurological/behavioral terms rendering unnecessary the need for a priori mentalist categories.

The marriage of behaviorism with the neurosciences in their collective rejection for non-verifiable, mentalist assumptions of the mind created an environment conducive for AI, which in turn is largely based on the mechanical simulations of thought-processes in routine input-output procedures. Sir John Eccles, a Nobel prize winning neuroscientist earlier on in his career declared: "We can, in principle, explain all...input-output performance in terms of activity of neuronal circuits; and consequently, consciousness seems to be absolutely unnecessary!....as neurophysiologists we simply have no use for consciousness in our attempts to explain how the nervous system works"[22]

The reinterpretation of mind in terms of neurology and neurology alone (or mechanism and mechanism alone) has provided a rhetorical background between the vitalists (who would see such a move purely as reductionism) and the materialists (who see such a move as an act of representation). What does all this mean rhetorically? The act of restating the mental states or the residues of one's interior consciousness in chiefly materialist terms implies at least two things: translation and transduction. Translation occurs when a neuroscientist chooses a particular set of mental variables, as the governing lens with which to see the manifestation of intelligent behavior and then restates these phenomena into tangible, operational terms. 'Computation' through a "set of instruction tables" is one of the key operational terms in Turing's thought. The operational terms in symbolic AI would include key ideas such as the retrieval of memory from a "stored symbolic database," "problem solving as logical inference," "cognition as centralized," "the environment as a problem domain," and the body as an "input device."[23] The operational terms in connectionist AI (artificial neural network) would include key ideas such as: "memory as pattern re-creation, problem solving as pattern completion and pattern transformation, and cognition as increasingly decentralized."[24]

Transduction occurs when one transfers or converts energy into another form. In our bodies, transduction occurs when physical energy is converted into an assortment of nervous signals. If one were to offer a neurological account of a simple activity such as the retrieval of information, one would have to pay careful attention to the conjunction of mental nodes at certain neural spaces where specific information pertaining to color or so forth is processed. An act of transduction occurs, in that, let us say in remembering a brilliant visual landscape - a flurry of brilliant colors are converted into a series of electrical signals, and retrieved from the storage area in that particular region of the brain. As A. Damasio and H. Damasio explain:

> ...the level at which knowledge is retrieved (e.g. superordinate, basic objects, subordinate) would depend on the scope of multiregional activation. In turn, it would depend on the level of convergence zone that is activated. Low level convergence zones bind signals relative to entity categories....Higher level convergence zones bind signals relative to more complex combinations...The convergence zone capable of binding entities into events...are located at the top of the hierarchical streams, in anteriormost temporal and frontal regions.[25]

In simpler words, the authors are referring to distinct and overlapping neural systems that offer access to different types of knowledge, i.e. analytical, spatial etc. If the nature of the information being processed is more complex, the higher the degree of coordinate activity is needed. Clark suggests, "higher-level capacities (such as grasping concepts) are, however, depicted as depending on the activity of multiple...areas (in sensory and motor cortices) mediated by the activity of multiple convergence zones."[26] One could even say that transduction takes place at a dialogic level, since there is a concurrent attempt to look at information as an interplay of interweaving neural systems, and by the same token, look at the neural pathways as the embodiment of coordinated knowledge of some kind or the other.

The dominance of such rhetorical strategies where the whole is replaced by the sum of additive parts in quite common in neuroscientific and AI circles.

Logical and Mathematical Origins

The relationship between philosophy and the sciences (including mathematics) has been a peculiar one. In the twentieth century two prominent rival positions espousing the interrelationships or lack thereof have emerged, namely the Russellian and the Wittgensteinian.[27] Bertrand Russell suggested that philosophy should seek its validation from the sciences by chiefly relying on scientific induction.[28] Both philosophy and science were engaged in the pursuit of knowledge insofar as the generation of theories and hypotheses were concerned. In Russell's scientific worldview, philosophy should take on the role of facilitating the evolution of science – however, a fine distinction was made between the two as follows:

> To a great extent, the uncertainty of philosophy is more apparent than real: those questions which are already capable of definite answers are placed in the sciences, while those only to which, at present, no definite answers can be given, remain to form the residue which is called philosophy.[29]

Ludwig Wittgenstein, on the other hand, advocated a disjunction between philosophy and the sciences. The province of philosophy is to clarify the nature of a concept using logic. The crux of Wittgenstein's argument is elucidated with axiomatic certainty, in that the source of a "philosophical problem often lies in the a crucial and often elusive difference between the *surface grammar* of a concept-word and its *depth* or *logical grammar,* or in the philosopher's tendency to treat what are disguised *grammatical* *propositions* as if they were empirical propositions."[30] Wittgenstein eschewed the idea that philosophers produce new knowledge:

> You always hear people say that philosophy makes no progress and that the same philosophical problems which were already preoccupying the Greeks are still troubling us today. I read: '...philosophers are no nearer to the meaning of "Reality" than Plato got....How extraordinary that Plato could have even got as far as he did! Or that we could not get any further![31]

In other words, Wittgenstein seems to suggest that the study of philosophy should consist of clarifying concepts and theories and not to draw inductive generalizations or to formulate theses. In his 1930 lectures, he declared:

> What we find out in philosophy is trivial; it does not teach *us new facts*, only science does that. But the proper synopsis of these trivialities is enormously difficult, and has immense importance. Philosophy is in fact the synopsis of trivialities.[32]

Wittgenstein almost seems neo-Platonic in his preoccupation with "ideas," which is a stark contrast to the Aristotelian empirical worldview that sowed the seeds for the development of scientific induction as a methodology and as a practice. In *Philosophical Investigations*, the Wittgensteinian project is summarized as follows:

> It was true to say that our considerations could not be scientific ones. It was not of any possible interest to us to find out empirically...there must not be anything hypothetical in our considerations. We must do away with all *explanation*, and description alone must take its place. And this description gets its light, that is to say, its purpose, from the philosophical problems. These are, of course, not empirical problems; they are solved, rather, by looking into the workings of our language, and that in such a way as to make us recognize those outworkings: *in spite of* any urge to misunderstand them. The problems are not solved, by giving new information, but by arranging what we have always known. Philosophy is a battle against the bewitchment of our intelligence by means of language.[33]

Wittgenstein who declared that philosophy does not teach us new facts, would probably approach the philosophy of mind not so much as an empirical enterprise to investigate the mechanism of the mind, but more as a linguistic enterprise whereby one has to clarify concepts such as "intelligence," "consciousness" and so forth.

Turing, although primarily a mathematician and logician, was not impervious to the disagreements between Russell and Wittgenstein and personally grappled with philosophical questions regarding the nature of the mind. Russell's position that philosophy must base itself on science is rejuvenated in Turing's philosophy of mind, in that Turing moves away from metaphysics with its emphasis on ontology and instead concurred to the view that a proper conception of mind could only be gained from studying the brain sciences. Turing hypothesized that the cognitive processes of the brain are machinelike, and attempted to draw isomorphic parallels in conceptualizing the development of machine intelligence. The conception and development of machine intelligence thus required a rich cross-fertilization of disciplines such as the cognitive sciences, mathematics, philosophy, electrical engineering and so forth.

Turing's conception of a universal Turing machine emerged as an off-shoot of his attempt to respond to Hilbert's metamathematical challenge of proving the internal consistency, compatibility and solvability of every mathematical statement formulation. Among the important questions that Hilbert raised dealing with the completeness, the consistency and decidability of mathematical propositions spawned path-breaking work in mathematical logic that carved open a space to provide a logical account of computational theory. In 1900 at the Second International Congress of Mathematicians Hilbert threw the gauntlet to his fellow mathematicians in his famous keynote address:

> If we would obtain an idea of the probable development of mathematical knowledge in the immediate future, we must let the unsettled questions pass before our minds and look over the problems which the science of today sets and whose solution we expect from the future....However unapproachable these problems may seem to us and however helpless we stand before them, we have, nevertheless, the firm conviction that their solution must follow by a finite number of purely logical processes...*This conviction of the solvability of every mathematical problem is a powerful incentive to the worker. We hear within us the perpetual call: There is the problem. Seek its solution. You can find it by reason, for in mathematics there is no ignorabimus.*[34]

There are three prominent responses to Hilbert's challenge to the mathematical world that are pertinent to the history of computing. Russell's attempt to articulate a decisive response to Hilbert's challenge led him to the now famous Russell's paradox, which essentially is a paradox of sets that foreshadowed Russell's theory of types. For instance, if one were to consider A which is defined as the set containing all sets that are not members of themselves one would on further reflection pose the inevitable question 'does set A contain itself?' For instance, if one were to consider A which is defined as the set to contain all sets that are not members of themselves one would on further reflection pose the inevitable question 'does set A contain itself?' If the answer is yes, one should say that set A contains itself. However, when the answer is in the affirmative, the basic defining condition that set A should not belong to set A is violated, leading one to a contradiction. If the answer is no, one should say that set A does not contain itself. The second response also violates the basic defining condition namely that if set A does not belong to itself, then it would belong to set A.

Russell's paradox delivered a mortal blow to Frege's project whose aim was to provide a cohesive system whereby all arithmetical notions were definable within a certain logical system and that all theorems of arithmetic were in turn theorems of the system. Frege claimed that an expression such as *f(a)* could be simultaneously expressed as a function of the argument *f* and a function of the argument *a*. Russell's discovery of the antimony that overshadowed Frege's contributions is best expressed as follows:

> this view (that *f(a)* may be viewed as a function of either *f* or of *a*) seems doubtful to me because of the following contradiction. Let *w* be the predicate: to be a predicate that cannot be predicated of itself. Can *w* be predicated of itself? From each answer its opposite follows. Likewise there is no class (as a totality) of those classes which, each taken as a totality, do not belong to themselves. From this I conclude that under certain circumstances a definable collection does not form a totality.[35]

Russell solves this logical conundrum with an elegant postulation of a logic machine. A logic machine essentially is premised on the concept of a logical transformation as an operation that is carried out over a quantum of time. Russell formulated a set of logical operations in which a particular problem can be expressed as a 'program' with a specific set of operations to flow. The program can be turned on in such a manner that every logical transformation or inference is implemented, thereby when the process is completed, one has a definite answer that circumvents the hitherto unresolvable paradox. If one were to apply the concept of the logic machine to the problem of set A – the following would happen. At one point, the answer will be yes and the program will continue running, and after a certain point the answer will be no. The program will continue in an infinite loop alternating between yes and no. The basic idea entailed ordering the sentences of a language or theory into a hierarchy (starting with sentences about individuals at the bottom level, sentences about sets of individuals at the next level, sentences about sets of sets and so forth) such that one can avoid reference to a set as the set of all sets, without being self-constrained within the concept of a 'set.' Along with Alfred N. Whitehead in *Principia Mathematica*, Russell issued a declarative that "whatever involves all of a collection must not (itself) be one of the collection."[36] Russell's conception of the logic machine along with the Boolean binaries of 0 and 1 assisted Turing with an operational framework to conceptualize his theoretical prototype for the computer.

The second response to Hilbert's program that almost created the demise of Hilbert's foundationalist certainty is the now renowned Gödel's incompleteness theorem. It is worth mentioning that Gödel has been misread by many as promoting a form of mathematical and epistemological relativism, the best example being Dale Cyphert's otherwise brilliant monograph "Strategic Use of the Unsayable" published in the *Quarterly Journal of Speech.*[37] Gödel's incompleteness theorem dealt with the recursive function of self-referentiality that made the quest for absolute provability of all mathematical axioms at all points of time impossible, however this has mistakenly led many to rush headlong in writing a rather precipitous and often bitter epitaph for the queen of the sciences.

Gödel's contribution in effect, opened up a space for dealing with the open-endedness inherent in mathematical and formal systems despite valiant efforts to foreclose and eliminate self-referentiality. Completeness, in mathematical parlance, chiefly refers to the property that every valid axiom or formula of a particular formal system is provable within that system. Gödel in his 1931 paper entitled "On Formally Undecidable Propositions of Principia Mathematica and Related Systems" posits the view that there exists an undecidable proposition (that neither it nor its negation is provable within the system) within every system: "We now construct an undecidable proposition of the system, *P(rincipia) M(athematica)*, that is, a proposition *A*, for which neither *A* nor not *–A* is provable…"[38] The undecidable proposition otherwise known as the first incompleteness theorem (G1), is followed by a second incompleteness theorem (G2) that is directly relevant to Hilbert's program. The second incompleteness theorem provides a philosophical justification for the implausibility of Hilbert's program, although Gödel himself explicitly states no intentions of abandoning Hilbert's formalistic viewpoint. The second incompleteness theorem holds that any logical statement which expresses the consistency of a system and which can be represented as a formula within the system is in itself among the formulas not provable in the system. There are three significant corollaries to the second incompleteness theorem:

> first, that any consistency proof for a theory, *T*, of which G2 holds will have to rely upon methods more logically powerful than those of theory *T* itself; second, that (in any significant case) a consistency proof for theory *T* can yield no epistemological gain and so cannot provide a satisfactory answer to the skeptic regarding *T's* consistency; and third, that as a result of this, G2 is not

> strictly implying the outright failure of Hilbert's program, at the very least indicates that modifications are to be made.[39]

Hilbert's program had two distinctive goals regarding the foundations of mathematics, descriptive and justificatory in nature. The descriptive goal entailed the complete formalization of mathematics. The justificatory goal was epistemological in nature, in that it necessitated the discovery of a finitary proof of those essential non-finitary aspects of mathematics. Gödel demonstrated that mathematics was incomplete since there existed assertions which could neither be proved nor disproved.

Extrapolating these insights, one could write arithmetical statements that referred to themselves, not unlike the Epimenides paradox (All Cretans are liars. Epimenides is a Cretan. Was Epimenides lying when he declared, "all Cretans are liars"?). Gödel's conclusion that there are undecidable propositions even within a formalized mathematical system disabused the certainty of Hilbert's challenge. If one were to rely chiefly on formal logic alone, a complete formalization of the axioms on which mathematics is based would be necessary. Under these circumstances, Gödel would respond that regardless of what the axioms are, there would always be within a system propositions that are decided by intuitive means. What this really means is that "formal logic cannot provide an ultimate criterion of validity of mathematical assertions."[40] Gödel, on one hand demonstrated the aspect of incompleteness however, he did not go so far as to ascertaining a way in which one could define what set of questions were decidable or not.

Turing approached Hilbert's metamathematical challenge in mechanical terms, meaning that a mechanical process might perhaps hold the answer.

Turing's Idea of Thought

Turing's hypothesis of a universal Turing machine can be seen not only as a blueprint for the future development of machine intelligence, but also sheds light on his assumptions of mind. The Turing machine comprises an infinite tape (on which could be written any series of two symbols 0 and 1) and a computational unit. The computational unit contains a program that follows a series of commands made up from a list of operations that Turing specified:

. Read Tape
. Move tape left
. Move tape right
. Write 0 on the tape
. Write 1 on the tape
. Jump to another command
. Halt

Pertaining to the immediate theoretical task at hand, Turing's paper on computable numbers reported the presence of unsolvable problems by extending Cantor's diagonal argument. Cantor suggested that rational numbers could give rise to irrational numbers – if one were to consider the rational numbers or ratios between 0 and 1, the list would be as follows: ½, 1/3, ¼, 2/3, 1/5, 1/6, 2/5, ¾, 1/7, 3/5, 1/8, 2/7, 4/5, 1/9, 3/7, 1/10. The next step is to convert the ratios into infinite decimals. If one were to consider the diagonal number .5306060020040180...and then change each digit by adding each by 1 except by changing a 9 to a 0, this would give rise to an infinite decimal beginning .6417171131151291...which happens to be an irrational number.[41] If one were to extrapolate the analogy that rational numbers could give rise to irrational numbers, one could say that computable numbers (a term that Turing used to refer to real numbers which could be calculated by a machine using a definite set of rules) could give rise to noncomputable numbers by means of a diagonal argument. Turing's discovery of non-computable numbers dealt a decisive blow to Hilbert's project.

However, on the brighter side, it spawned the new field of computational theory and eventually the birth of the modern computer. Turing employed the notion of "unsolvable" problems to draw striking parallels between the workings of the human brain and the computer. If there are problems that were unsolvable for a machine that follows fixed laws, the same is true for humans, essentially biological machines following 'natural' laws. Turing conceived the digital computer to be a counterpart to the human brain; in other words, the brain is considered a discrete state machine that processes electrical impulses and specific neural inputs, not unlike the machine whose magnetic tape processes information in terms of the symbols 0 and 1. Although the human brain is a biochemical entity, looking for isomorphic resemblances with the computer and vice-versa is based on the premise that humans are biological machines and that the basis for the isomorphism is articulated because both the entities process information through electrical impulses. An interesting fact worth noting is that the term 'computer' in the past was used to describe a human engaged in the act of computing – this usage is no longer in vogue.

Turing described the information-processing abilities of the human computer primarily in terms of symbol-manipulation, using a similar language with which he described the universal Turing machine:

> The behavior of the computer (in this case, Turing was referring to a person doing calculations) at any moment is determined by the symbols which he is observing, and his 'state of mind' at any moment....We will also suppose that the number of the states of mind which need to be taken into account is finite. The reasons for this are the same character as those, which restrict the number of symbols. If we admitted an infinity of states of mind, some of them will be 'arbitrarily close' and will be confused. Again, the restriction is one which seriously affects computation, since the use of more complicated states of mind can be avoided by writing more symbols on tape.[42]

Turing considers human thought to be a step-by-step process entailing a series of operations, a form of symbol-manipulation that engenders the various 'states of mind.' Although human memory seems to be limited, Turing implicates human thinking with the same descriptive procedures employed for the digital computer. The Cartesian argument of substance dualism finds itself replaced by a robust physicalist viewpoint, charting out the topography of the mind in terms of material nodes and pathways. Through the use of discrete symbols, Turing provides a material basis for the mind. In Turing's words:

> Let us imagine the operations performed by the computer to be split up into 'simple operations'which are so elementary that it is not easy to imagine them further divided. Every such operation consists of some change in the physical system consisting of the computer and his tape. We know the state of the system if we know the sequence of the symbols on the tape, which of these are observed by the computer (with a special order), and the state of the mind of the computer.[43]

The human computer is a discrete state machine processing distinct symbols. The *machina* in Burkean parlance emerges as the terministic screen or coloring lens whereby Turing perceives human cognition. By the same token, Turing uses the complex dynamics of human cognition as a conceptual framework to guide the evolution of machines. Turing propounded three theses to speak about human-machine cognition:

> Programming could be done in symbolic logic and would then require the construction of appropriate interpreters,
> Machine learning is needed so that computers can discover new knowledge inductively from experience as well as deductively,
> humanized interfaces are required to enable machines to adapt to people, so as to acquire knowledge tutorially.[44]

Envisioning a human model for machine-learning, Turing envisaged the machine to add its distinctive stamp of originality, and furthermore the machine would serve as a stimulus in enhancing human learning due to the common body of symbols both machines and humans use in cognitive processing. Digital computing machines "will eventually stimulate a considerable interest in *symbolic logic* and *mathematical philosophy* (emphasis mine)," and the "language in which one communicates with these machines" will be "the language of instruction tables" which is essentially a form of "symbolic logic."[45] Turing proposed a process-oriented model of machine learning, whereby the machine can internalize, adapt and later on modify the instruction:

> Let us suppose we have set up a machine with certain instruction tables, so constructed that these tables might on occasion, if good reasons modify these tables. One can imagine that after the machine has been operating for some time, the instructions would have altered out of all recognition, but nevertheless still be such that one would have to admit that the machine was still doing very worthwhile calculations. Possibly it might be getting results of the type desired when the machine was first set up, but in a much more efficient manner. In such a case one would have to admit that the progress of the machine had not been foreseen when its original instructions were put in. It would be like a

> pupil who had learnt much from his master, but had added much more by his own work[46]

In the same lecture, Turing asserts, "what we want is a machine that can learn from experience."[47] Turing makes an eloquent case for creative machine learning by going beyond the rote-capacities that are normally attributed to them. The crux of Turing's argument rests on the concept of "storage." The concept of storage is in turn intertwined with the concept of memory, which is a fundamental attribute of human existence. The "real life" events of the past, although receding rapidly, from our foregrounded consciousness, remains alive in the form of memories. Humans learn by association such that when a similar event happens in the present or in the future, we readily fall back on our frames of experiences that provide us with a reference point to make sense of what we are experiencing now. This could be true of something as simple as arithmetic or something as complex as abstract philosophy. Turing's central argument for the flourishing of artificial intelligence is based on the concept of memory, especially memory when enhanced exponentially compared to earlier machines:

> It might be argued that there is a fundamental contradiction in the idea of a machine with intelligence. It is certainly true that 'acting like a machine,' has become synonymous with lack of adaptability. But the reason for this is obvious. Machines in the past have had very little storage, and there has been no question of the machine having any discretion.[48]

The type of memory that Turing appears to elucidate is an associative memory relying on behaviorist models of stimuli-response and not on any transcendental or metaphysical intimations of consciousness. Memory and learning are considered purely as biochemical and computable processes that are calculable by specific algorithmic procedures.

Turing sublates the distinction between intelligence and consciousness, by circumventing ontology and instead resorting to the emerging or perhaps reigning philosophy of mind viz. behaviorism. Turing reformulated the conception of intelligence as being confined to distinct biological entities and posited the idea that intelligence could emerge out of physical symbol systems with the appropriate neural chemistry or circuitry to perform tasks that would otherwise intelligence when performed by humans. Turing compares the infant automaton with that of the human mathematician, and suggests that machines also go through a similar learning curve not unlike humans. Human mathematicians are prone to error and so are machines. However, machines tend to be discounted more easily than humans, hence "fair play" becomes vital to furthering the ascent of intelligent machines:

> To continue my plea for 'fair play for the machines' when testing their I.Q. A human mathematician has always undergone an extensive training. This training may be regarded as not unlike putting instruction tables into a machine. One must therefore not expect a machine to do a very great deal of building up of instruction tables on its own. No man adds very much to the body of knowledge, why should we expect more of a machine? Putting the same point differently, the machine must be allowed to have contact with human beings in order that it may adapt itself to their standards. The game of chess may perhaps be rather suitable for this purpose, as the moves of the machine's opponent will automatically provide this contact.[49]

Turing's predictions insofar as chess is concerned have amply been fulfilled. IBM's *Deep Blue* only a couple years ago beat the highest rated human chess player. Technically speaking, the best chess player in the world is neither Kasparov nor

Kramnik but a machine. It is common knowledge that some chess tournaments with a Grand Master rating are routinely won by machines. But the ultimate test for machine cognition even according to Turing, is not mathematical (Chess can be seen as a closed and formal mathematical system with an *n* number of permutations and combinations) but verbal. A rhetorical analysis using Burke's terministic screen of Turing's "imitation game" will be the focus of the next section.

Thought as Computation and Computation as Metaphor

The development of the modern digital computer has much to do with the development of mathematical logic. What is perhaps noteworthy is mathematical logic's insistence on a finite number of steps to accomplish a particular proof or so forth. In von Neumann's words:

> Throughout all modern logic, the only thing that is important is whether a result can be achieved in a finite number of elementary steps or not. The size of the number of steps which are required, on the other hand, is hardly ever a concern of formal logic. Any finite sequence of correct steps is, as a matter of principle, as good as any other...In dealing with automata, this statement must be significantly modified. In the case of an automaton the thing which matters is not only whether it can reach a certain result in a finite number of steps at all but also how many such steps are needed.[50]

von Neumann's observation can be extrapolated to what Turing conceived as the thinking process. Thought is seen as computation, and thought processes are instantiations of program like procedures. In Turing's words:

> The idea behind digital computers may be explained by saying that these machines are intended to carry out any operations which could be done by a human computer. The human computer is supposed to be following fixed rules; he has no authority to deviate from them in any detail. We may suppose that these rules are supplied in a book which is altered whenever he is put on to a new job.[51]

The digital computer has three parts: "a) Store, b) Executive unit and c) Control."[51]

> The store is a store of information, and corresponds to the human computer's paper, whether this is the paper on which he does his calculations or that on which his book of rules is printed. Insofar as the human computer does his calculations does calculations in his head, a part of the store will correspond to his memory.[52]
>
> The executive unit is the part which carries out the individual operations involved in a calculation. What these individual operations are will vary from machine to machine.[53]
>
> We have mentioned that the "book of rules" supplied to the computer is replaced in the machine by a part of the store. It is then called the "table of instructions" It is the duty of the control to see that these instructions are obeyed correctly and in the right order. The control is so constructed that this necessarily happens.[54]

Turing does not provide an operational definition of intelligence or thinking, instead he argues and rightly so, that computation is largely an algorithmic process following a finite set of steps. The idea of a store, an executive unit and a control which the digital computer should have do not necessarily have direct one-to-one correspondences in Turing's conception of mind, yet he believed that human thinking was descriptive and prescriptive, in that, it followed a set of instruction tables from a 'book of rules:'

> If one wants to make a machine mimic the behavior of the human computer in some complex operation one has to ask him how it is done, and then translate the answer into the form of an instruction table. Constructing instruction tables is usually described as "programming." To program a machine to carry out the operation A" means to put the appropriate instruction table into the machine so that it will do A.[55]

In such a case, thinking is a computable process enacted through a series of discrete states with inputs and outputs. Digital computers are considered discrete state machines:

> Digital computers fall within the class of discrete state machines. But the number of states of which such a machine is usually enormously large...The computer includes a store corresponding to the paper used by a human computer. It must be possible to write into the store any one of the combinations of symbols which might have been written on the paper...Given the table corresponding to a discrete state machine, it is possible to predict what it will do. There is no reason why this calculation should not be carried out by means of a digital computer. Provided it will be carried out sufficiently quickly the digital computer could mimic the behavior of any discrete state machine.[56]

Ajit Narayanan remarks that underlying Turing's proposal are the following assumptions:

> Humans and a certain type of computer are of the same type and of the same type and so have the same mental predicates ascribable to them. Hence, all the mental predicates ascribable to humans are ascribable to computers of that type, and vice-versa.
> There is some overlap between humans and a certain type of computer which allows some mental predicates to be ascribed to both.
> Certain predicates can be equally ascribed to entities belonging to a variety of types (including both humans and computers) but these predicates do not assume full person status on the part of that to which they are ascribed.
> Finally, certain types of predicate can only be ascribed to certain entities, i.e. there is no overlap between humans and any type of computer.[57]

Turing did not oversimplify by any means human thinking, but believed that much of human thinking including verbal abilities belonged to the realm of computation. Not unlike the human machine, the digital machine may also have physical limits – however, its limitations should not be misinterpreted as an inability to engage in thinking.

The expansion of the capabilities of a formal system and importing its vocabulary by purportedly stating that the thinking capacities of humans (in whose case, conceptions such as 'thinking' and 'intelligence' tend to be less sharply defined) whose proclivity to engage in nonmonotonic reasoning and informal logic is well-illustrated by life experiences is a good example of an ampliative argument. The juxtaposition of thought with computation implies that all thought at some level or the other involves some kind of computation because it difficult to arbitrarily and a priori foreclose the boundary by saying aspect A of human experience is computable whereas aspect B of human experience is noncomputable, at least from the perspective of computational psychology.

By using the metaphor of 'computation' from digital computer in describing human behavior, Turing actually maps the language of artificial systems onto what he perceives are natural information-processing systems. The process could be reverse as well. This position is well-exemplified by David Marr (speaking of human vision) who argued that computational theories of behavior should be able to describe systems at three levels: a) the system's hardware in which processes are viewed as implementations of algorithms and data structures; b) algorithms and data structures are viewed as abstractions and c) the computations effected by the algorithms being executed, are seen as transformations of inputs to outputs - These levels from a natural processing system are then mapped onto an artificial system in terms of a) the physical processes of the hardware, b) the description of the virtual machine defined by a set of processes such as a target domain and c) the computational behavior of the virtual machine which carries out the task of the target domain.[58] Again, as stated earlier, it is not necessary to assume exact one-to-one correspondences between the store, an executive unit and a control from the digital computer and the human brain, although some computational psychologists have attempted to do so. However, what is salient is that using the metaphor of computation and describing human thinking primarily in terms of computation can generate an entire conceptual vocabulary (this will be addressed in greater detail in the next chapter, where we will talk about how the 'mind is a computer' has generated an entire vocabulary for computational psychologists).

The computational metaphor also invokes the idea that something can be shown or demonstrated. When something is shown, there is a sense that something is being specified. However, when something is being specified, there is also a sense that something else is being unspecified. When what is hitherto unspecified is attempted to be specified, there is a sense that there is something else that is still being unspecified. Therefore, although one could theoretically latch on to the dream of ultimate formalizability although neither Gödel nor Turing subscribed to this view – there will always remain elements that are nonformalizable rendering the view that thought is equivalent to computation only partially accurate.

Imitation Game

Kenneth Burke made a significant distinction between "scientistic" and "dramatistic" approaches to language.[59] A scientistic conception of language pertains to what is and what is not, the dramatistic conception of language pertains to the realm of symbolic nature of language, and is "exercised about the necessarily *suasive* nature of even the most unemotional scientific nomenclatures."[60] Even in the act of drawing the distinction, Burke develops the notion of a terministic screen that ironically blurs the distinction between the two. Even the most unemotional language based on pure logic entails a particular epistemic orientation towards the nature of the things, and promotes a particular way of knowing at the expense of other ways of knowing. A terministic screen as a rhetorical lens entails any symbolic use of language within which are embedded are three elements: a) a reflection of reality, b) a selection of reality and c) a deflection of reality. Any insight by virtue of its selection also entails an act of omission, since there is a directing of attention: "any nomenclature necessarily directs the attention into some channels rather than others."[61] A further extrapolation of the terministic screen as a conceptual lens sheds light on how one scrupulously selects and directs certain types of signification more than others.

Turing proposed the imitation game as a formal test for machine intelligence. By replacing the question "can machines think?" with "are there imaginable digital computers which would do well in the imitation game?"[62] Turing replaces the actual conception of thinking with acting like thinking. The imitation game, as is well known, is a variation of the parlor game played originally by a man, a woman and an interrogator is now played by a computer, a human and a human interrogator. The interrogator types in a series of questions and attempts to carry on a regular conversation, during the course of which the goal is to make an explicit distinction between the human and the computer. If the computer's simulation of mental phenomena (verbal phenomena) is adequately sophisticated to pass for a human, the machine has passed the Turing test.

Turing rejects outright wrestling with the philosophical dimensions of concepts such as 'intelligence' or 'thinking' because it might invoke ambiguity, and opts for a purported unambiguous engagement of those terms purely in pragmatic terms:

> If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous terms.[63]

Perhaps a little more explication is necessary to understand the rhetorical significance of Turing's strategy – it is common knowledge that most cognitive scientists are interested in experimental evidence, models and theories evidenced from mental behavior and if this were the case, emphasis can only be placed on manipulation of tokens or symbols or overt mental behavior since the minds are considered inaccessible from a purely neurological perspective. Daniel Dennett remarks are germane to this line of reasoning:

> The cognitive scientists marshals experimental evidence, models, and theories to show that people are engaged in surprisingly sophisticated reasoning processes of which they can give no introspective account at all. Not only are minds accessible to outsiders; some mental activities are more accessible to outsiders than to the very 'owners' of those minds.[64]

Implicit in Turing's argument is a rejection of a wide gamut of accounts including the intuitionist, reflective and introspective largely because of the unstated presupposition that there is no access to the mind or internal operations, without specifying a domain of mental activity that can be simulated. Turing's willingness to include verbal intelligence, hitherto considered unique and *sui generis* to human systems, is perhaps that conversational abilities are programmable and the open-endedness of normal, human conversational settings can be reinterpreted and foreclosed in algorithmic operations.

The idea of an imitation game is also seen as leveling the playing ground:

> We do not wish to penalize the machine for its inability to shine in beauty competitions, nor to penalize a man for losing in a race against an airplane. The conditions of our game make these disabilities irrelevant... The game may perhaps be criticized on the grounds that the odds are weighted against the machine. If the man were to try and pretend to be the machine he would clearly make a very poor showing. He would not be given away at once by slowness and inaccuracy in arithmetic. May not machines carry out something which ought to be described as thinking but which is very different from what a man does? This objection is a very strong one, but at least we can say that if nevertheless, a machine can be constructed to play the imitation game satisfactorily, we need not be troubled by this objection...and it will be assumed that the best strategy is to try to provide answers that would naturally be given by a man.[65]

If the machine provides the type of answers that would be naturally provided by humans (that is essentially acting like humans) it is said to be intelligent or said to be a thinking entity. Turing's certainty that acting like thinking and thinking are interchangeable is unmistakable:

> The original question, "Can machines think?" I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted. I believe further that no useful purpose is served by concealing these beliefs.[66]

The incontrovertible argument that Turing provides is still disputable, however, the replacement of the original question 'can machines think?' is carried out in earnest in the cognitive sciences with primary emphasis being placed on manifest mental behavior or the acting out of those mental behavior. Therefore, the replacement of 'thinking' with 'acting like thinking' in of itself becomes a terministic screen, a "conjecture" of "great importance" that "suggest useful lines of research."[67] What the implications are will be discussed in the following paragraphs.

The obvious advantage based on Turing's selection of the schema of 'producing intelligent behavior' is that one could go about constructing machines that play chess, checkers, engage in pattern-recognition, speech recognition among a range of other activities that would require mental activity when produced by humans. More specifically, at this juncture, the Turing test is about the ability to carry on a conversation. If a group of astute programmers are able to envision a diverse assortment of conversational scripts that an average person is likely to engage in – after gleaning social scientific evidence on topics that are recurring among conversants and also are able to script out particular 'intelligible' responses to a scenario where a computer has no contextual reference to – it is possible to say that at some level, it is possible to trick the interrogator into a difficult situation in not being to able to discriminate the responses between a human and a computer. A tentative example of such a script may be:

Interrogator: I am suffering from a severe case of pre-senile dementia. Have you any suggestions?

Computer (assuming that it has no background knowledge of pre-senile dementia, may very well respond): I am really sorry. Have you talked to a psychiatrist?

For instance, the programmer may train the computer program to associate 'dementia' with 'psychiatrist' and 'suffering' with a response 'I am sorry,' but this does not tell us whether the computer understands what the interrogator is describing. Therefore, the idea of producing intelligent behavior that essentially revolves around the notion of acting like it is thinking in itself is not a strong enough criterion for ascertaining computational intelligence. It does not follow that the above-mentioned statement precludes the possibility of conceiving machines as thinking entities, but 'intelligent' behavior in of itself tells us little about the nature of intelligence. J. R. Lucas argues the point that 'acting' or practice in itself tells us little about intelligence rather lucidly:

> In asking the mechanist rather than the machine, we are making use of the fact that the issue is one of principle, not of practice. The mechanist is not putting forward actual machines which actually represent some human being's intellectual output, but is claiming instead that there could in principle be such a machine. He is inviting us to make an intellectual leap, extrapolating from various scientific theories and skating over many difficulties. He is quite entitled to do so. But having done this he is not entitled to be coy about his in-principle machine's intellectual capabilities or to refuse to answer embarrassing questions. The thought-out experiment, once undertaken, must be thought through.[68]

Turing's response to the above-mentioned critique may be along the lines – 'after all, we humans are programmed machines, so why hold the computer to a harsher test.'

Turing argues that the type of originality we attribute to humans could be attributed to machines as well:

> Let us suppose we have set up a machine with certain instruction tables, so constructed that these tables might on occasion, if good reason arose, modify those tables. One can imagine that after the machine has been operating for some time, the instructions would have altered out of all recognition, but nevertheless would still be such that one would have to admit that the machine was doing very worthwhile calculations. Possibly it might still be getting results of the type desired when the machine was first set up, but in a more efficient way. In such a case one would have to admit that the progress of the machine has not been foreseen when its original instructions were put in. It would be like a pupil who learnt much from his master, but had added much more by his own work. When this happens I feel that one is obliged to regard the machine as showing intelligence.[69]

We have not gone much further from the idea of 'showing intelligence,' however it is whole notion of showing intelligence that propels much of Turing's arguments for machine intelligence. Mathematical logic enables a programmer to represent both numerical and non-numerical content through a string of symbols. The style of programming entails a number of steps: 1) modularizing a program into pieces that are coherent, such as subject, predicate and so forth, 2) describing the structure of the program in a way that is transparent, 3) describing clearly and precisely both the data models and the data structures – and explaining the operation performed by each procedure, describing the complete procedure and explaining how the inputs relate to the outputs, 4) using intelligible names for procedures and variables, 5) avoiding the use of explicit constants – if one were referring to the number of balls in a room, refrain from using numbers such as 6 or 5 but instead use a constant such as *NumberOfBalls*, so that changes can be made whenever necessary and so forth. All of this only means that if one is attempting to perform a computation, one should state the algorithm with set of finite steps that are computable. The ability of the computer to follow the commands of the algorithm could result in producing intelligent behavior, but that in of itself as stated earlier does not tell us what intelligence is.

Turing's idea of 'showing intelligence' whether it is performing a calculation or carrying a conversation is a form of computation. So, the terministic screen acts out in the following fashion –Thought = computation and computation = algorithm procedures enacted on a universal Turing machine and the brain is a type of universal Turing machine. Since the computer can very easily perform the algorithms (not unlike the human who is also performing similar algorithms) Turing establishes a similitude both of operation and function, therefore the differences in 'hardware' between the mind and the digital computer are erased. Whether this similitude that Turing establishes is valid is certainly debatable, however, from a purely behaviorist perspective Turing is right. Under Stimuli-Response models, greater emphasis is placed on describing the process of mental behavior in terms of inputs and outputs, therefore the 'simulation' of thinking is considered 'thinking' itself.

By selecting the above-mentioned paradigm of intelligence, Turing reinforces behaviorist models of intelligence:

> We normally associate punishments and rewards with the teaching process. Some simple child-machines (learning machines) can be constructed or programmed on this sort of principle. The machine has to be so constructed that events which shortly preceded the occurrence of a punishment-signal are unlikely to be repeated, whereas a reward-signal increases the probability of repetition of the events which led up to it.[70]

But, on the other hand, it provides a tenuous conception of intelligence. As Aaron Sloman explains the description of computation does not by itself explain the act of thinking:

> The formal notion of computation, equally applicable to physical and non-physical mathematical structures, does not on its own enable us to build useful engines or explain human or animal behavior. An abstract instance of computation cannot make anything happen. This shows that being a computation in the formal sense is not a sufficient condition for being an intelligent, behaving system, even though the formal theory provides a useful conceptual framework for categorizing some behaving systems.[71]

So, what we really have is criteria that hints towards some framework of thinking, without actually providing an operational definition of thinking nor intelligence. Sloman further argues:

> If being governed by rules involves understanding the rules. Understanding is a part of what we wish to explain: we must not assume it as a primitive. Can we avoid circularity by using a notion of being 'governed' by rules that does not presuppose understanding, but might provide a basis for it? Computational processes would then be processes controlled by rules. But if rules are not understood, i.e. meanings play no role in the control, then the words 'rule' and 'governed' are misleading, and we are simply left with the notion of processes controlled by something. What, then, is control? Various forms of control – mechanical, hydraulic, electronic, chemical, etc have been studied by control theory. It seems acceptable to say that all computations are controlled processes, but are all controlled processes computational?...Control is just a special case of the notion of causation: one thing controls another if there is some sort of causal relation between the controller and the controlled...We now seem to be moving towards a 'computation' that is so general...that it includes everything, trivializing the claims that mental states are computational.[72]

Given the ambiguity in the description that thought is computational, what is left is merely Turing's claim that digital computers (and by the same token implicating human brains as well) have discrete states and the internal state at any moment is determined by the last state and the input signal.

Even while acknowledging Turing's mathematical genius in ushering in the computer revolution; one should not be impervious to the gaps in Turing's model of cognition (how the actual terministic screen plays out in generating the rhetoric will be discussed more in chapter V pertaining to symbolic logic, in order, to avoid repetition). However, Turing need not necessarily be blamed for these conceptual gaps because he was a product of his time; a time where behavioristic models of cognition prevailed.

The next section will look at Turing's preemptive counter-arguments against potential objections that might be raised against the 'imitation game' or 'machine intelligence' in general.

Turing's Response to Objections

A rhetorical reading of Turing's counterarguments is instructive in unpacking the epistemological assumptions of his discourse of mind. As Alan Gross explains, "from the point of view of rhetoric as critique, the rationality of science consists in the continuing dialectic among its legitimate reconstructions, each the surrogate for the informed assent of the interpretive community."[73] The continuing dialectic between Turing and his fellow mechanists also implies a subsequent rejection of non-mechanist arguments by exposing the inadequacy of the counterclaims. Turing thus simultaneously anticipates and refutes possible objections towards achieving computational intelligence, which were as follows: a) the theological objection, b) the "heads in the sand" objection, c) the mathematical objection, d) the argument from consciousness, e) the argument from various disabilities, f) Lady Lovelace's objections, g) argument from continuity in the nervous system, h) argument from informality of behavior and i) argument from extra-sensory perception.[74]

It has been tentatively suggested, especially in religious circles that one of the driving forces behind Cartesian substance dualism does not reside so much in Descartes' animosity towards 'matter' (because he does acknowledge that the body is a machine) but in his religious metaphysics where a disjunction between spirit and matter is made. The theological objection, which Turing encounters, is grounded on the premise that only humans possess souls, automatically disqualifying machines from the realm of intelligence because *thinking* is considered a function of the human's immortal soul. Before addressing the crux of Turing's counter-argument, it must be mentioned that many thinkers do believe in a soul yet it is an exceedingly difficult concept for scientific exploration. Having said so, Turing's offhand treatment of a metaphysical question that has perplexed many minds (all of whom have failed to provide a comprehensive or decisive treatment of this subject) is understandable. What is perhaps noteworthy is that Turing employs the same line of reasoning that he employs to attack the original question "Can machines think?" Due to the presence of a wide variety of interpretations vis-à-vis thinking, Turing finds the question innocuous and trivial at the same time. Therefore, it is not surprising that Turing asserts the diverging conceptions of the soul in various religious traditions and the lack of commonality thereof to his advantage.

Turing's response to the metaphysical question is one of straightforward rejection. Yet, in a seemingly conciliatory note to religionists, Turing adds that "in attempting to construct such machines we should not be irreverently usurping His power of creating souls, any more than we are in the procreation of children: rather we are, in either case, instruments of His will providing mansions for the souls that He creates."[75]

Turing makes a direct allusion to intelligent design, and declares that humans can also be causal agents in furthering the creative endeavor of the maker by making intelligent machines. On one hand, a person who has remote familiarity with Turing's own conversion from some form of Christianity or belief in an almighty to total atheism could boldly say that Turing was only being facetious in advancing the plausibility of the intelligent design argument. But on the other hand, Turing in spite of his nonchalant treatment of the subject places this objection as ordinal number one given his appreciation for the fact that most religionist or spiritualists would use the metaphysical counter-argument against the equation or reduction of human cognition to matter alone or the elevation of a machine to the status of being a person. One of the primary beliefs of Judeo-Christianity is the cardinal doctrine of *Imago Dei.* The radical shift from *Imago Dei* to *Imago Machina* might be theologically unbearable for the average religious person; therefore Turing out of graciousness instead of denying religious experience *per se* accommodates divine intervention by chastising believers for limiting the sovereignty of God. To a religious audience, one could read that Turing is suggesting (or indirectly admonishing) that one could not arbitrarily limit the providence of God in considering sentience or the state of having a soul to be 'specific' to humans.

However, to a non-religious audience (or an audience indifferent to metaphysical questions), Turing's response could be read between the lines. In short, to a religious person Turing's argument reads like a sincere exhortation for refraining from limiting the creativity of God or the manifest creativity of the creator, evidenced from the created making intelligent machines. And to a non-religious person, it could read as an attempt to merely avoid a theological question.

The 'head in the sands' objection is merely a fearful response to machine intelligence:

> The consequences of machine thinking would be too dreadful. Let us hope and believe that they cannot do so. But it affects most of us who think about it at all. We like to believe that Man is in some subtle way superior to the rest of creation. It is best if he can be shown to be *necessarily* superior, for then there is no danger of him losing his commanding position. The popularity of the theological argument is clearly connected with this feeling. It is likely to be quite strong in intellectual people, since they value the power of thinking more highly than others, and are more inclined to base their belief in the superiority of Man on this power.[76]

Turing is suggesting that the head in the sands objection is based on an anthropocentric bias and also on the grounds of illusory human superiority. While there may be merit to Turing's counter-argument, the substance of this critique may not necessarily apply to all opponents of machine intelligence. Some may oppose the notion purely on the notion that thinking is a product of living beings, while simulation may have more to do with artificial entities. Or to be more precise, the counterargument to Turing's refutation is to state that thinking can be construed as a product of biological systems.

The mathematical objection overplays the obvious limitations of discrete state machines: "the best known of these results is known as Gödel's theorem, and shows that in any sufficiently powerful logical system statements can be formulated which can neither be proved nor disproved within the system, unless possibly the system itself is inconsistent."[77] The inability to formalize all statements has led some to state that AI's models of cognition are not entirely accurate because of the presence of non-formalizable elements even within a formal system. Turing responds to this critique by stating "the short answer to this argument is that, although it is established that there are limitations to the power of any particular machine, it has only been stated, without any sort of proof, that no such limitation apply to the human intellect."[78] The unfairness of the mathematical objection is exposed when one considers the fact that humans are not subject to the same objection in spite of our obvious limitations. Turing is quite right in saying this point, however, what he fails to address is the argument that Gödel's incompleteness theorem dampens the epistemological confidence people might have in asserting that thinking, at least, human thinking is a mechanical process. The presence of nonformalizable elements implies that there are many aspects of thinking that do not neatly fall under the mechanist thesis which means that human thinking includes mechanistic plus nonformalizable elements as well which makes the assertion that the mind is just a computer debatable.

The argument from consciousness is best illustrated by Professor Jefferson's observation:

> Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain – that is, not only write it but know that it had written it. No mechanism could feel (and not merely artificially signal, and easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants.[79]

Turing refutes that position by stating that the "only way by which one could be sure that a machine thinks is to *be* the machine and to feel oneself thinking.... It is in fact the solipsist point of view. It may be the most logical view to hold but it makes communication of ideas difficult. A is liable to believe "A thinks but B does not" while B believes "B thinks but A does not."[80] While the argument against solipsism is powerful, what Turing fails to address is the aspect of creativity and lived experience – many songs, poems and so forth are written out of lived experience. When the Romantic poet Shelley remarked 'I fall on the thorns of life, I bleed' – he is not merely referring to a clever arrangement of syntax, but perhaps his own experience or the experience of other humans who suffer. Again, Turing speaks purely as a logician and not necessarily as a person who is a social commentator (but Turing himself is no foreigner to suffering, the experiences surrounding his person life and the rejection he received was very cruel).

The arguments from various disabilities suggest that there are certain things that machines cannot do. However, Turing responds that this argument can be refuted if the machine has adequate storage capacity. In other words, if a machine has adequate storage capacity it can display a diversity of behaviors and thus falsify the "argument from various disabilities." Purely from the perspective of mathematical logic alone, many things have not been accomplished such as pattern-recognition among others – it took a different way of approaching 'computing' namely neural networks to effect these type of activities. But to Turing's credit it might be said, that Turing next responds to Lady Lovelace's objection to Charles Babbage's Analytical Engine. Lovelace suggested that machines could only perform the tasks that we order it to do, and cannot do anything on its own. Turing responds by debunking the mystique of originality – that a human invention or idea always stems out of something else. Furthermore, machines can perform tasks with amazing speed and accuracy, that in Turing's words; they take us by "surprise."

The next objection is the "argument from continuity in the nervous system" which pertains to the fact that the nervous system is not a discrete state machine. To circumvent this disadvantage, Turing postulates the notion of a "differential analyzer" that can provide a range of values ranging from 3.12 to 3.16 with very close probabilities when asked the value of π. Thus if one were to play the imitation game, the differential analyzer can be indistinguishable from the human.

Turing rejects the "argument from informality of behavior" by deriding the anti-mechanistic view of human nature as being grounded on a false premise that humans are not regulated by "rules of conduct." Turing suggests that there are laws of behavior that suggest that humans are not unlike machines in that regard. Turing also rejects the "argument from extra sensory perception" by stating that computers can circumvent this difficulty if one were to incorporate a random number generator such that the computer can do things which defy the realm of ordinary science.

Scientific Ethos and the T-test

One of the key Aristotelian conceptions of public discourse is the notion of ethos. A literal translation from the Greek reads as character or credibility. Ethos can also be translated as a mask or a persona that the rhetor not unlike a thespian plays before his or her audience. Aristotelian ethos is based on three features: *phronêsis* (practical wisdom), *aretê* (virtue) and *eunoia* (good will).[81] The preceding section outlined a summary of Turing's kernel rebuttals of the nay-sayers. Turing appeals to pragmatic considerations in preferring behaviorism over ontological metaphysics, constantly invoking a seemingly ruthless examination of human intelligence to demystify any pretensions whatsoever that one might have of human transcendence, and ultimately extending a welcoming hand to machines to enter the hitherto exclusive club of intelligent species. Turing's recurring appeal to fair play enhances his ethos and the overall persuasive nature of his message.

The overall persuasive nature of a message corresponds directly to the extent to which it impacts the *pistis* (beliefs, convictions) of the audience. Ethos is best characterized by Aristotle as follows:

> (there is persuasion) through character whenever the speech is spoken in such a way as to make the speaker worthy of credence; for we believe fair-minded people to a greater extent and more quickly (than we do others) on all subjects in general and completely so in cases where there is no exact knowledge but room for doubt. And this should result from the speech, not from a previous opinion that the speaker is a certain kind of person; for it is not the case, as some of the technical writers propose in their treatment of art, that fair-

> mindedness (*epieikeia*) on the part of the speaker makes no contribution to persuasiveness; rather, character is almost so to speak, the controlling factor of persuasion.[82]

Aristotelian ethos does not refer to the intrinsic character of the rhetor, but pertains to the "character of the speaker constructed in the text, exhibiting fair-mindedness"[83] and adapting to the general character of the audience. It is safe to assume that Turing's message would have been lost were it not for the existence for a vibrant ideology of behaviorism and physicalism permeating the overall intellectual ambience of the times. As Willard Quine ruminates:

> ...the proposition that external things are ultimately to be known only through their action on our bodies should be taken as one of the coordinate truths, in physics and elsewhere, about initially unquestioned physical things. It qualifies the empirical meaning of our talk of physical things, while not questioning the reference. There remains abundant reason to inquire more closely into the empirical meaning or stimulating conditions of our talk of physical things, for we learn in this way about the scope of creative imagination in science; and such inquiry is none the worse for being conducted within the framework of those same physical acceptations.[84]

Ian Hacking traces the complicated roots of scientific epistemology, by grounding it in a tradition of referentiality whereby natural sciences demand heuristic and suasive power by drawing our attention to the fact that scientific knowledge is veridical and represents the world, while at the same time deflecting our attention from the eisegetical fact that scientific knowledge also intervenes in the world.[85] For every act of intervention, there is an essential conflation of the distinction between a theoretical model and a mirror. Burke's terministic screen although used religiously in this essay at the risk of exasperating the kind and gentle reader, is insightful in portraying the symbolic power of models or terminologies in rhetorically constructing a discursive community that pledges allegiance to the central tenets and doctrines of a particular epistemological orientation. In effect, there are two impulses – one attesting to the rhetorical invention of the rhetor constructing these terminologies: "In the unwritten cosmic constitution that lies behind all man-made Constitutions, it is decreed by the nature of things that each man is "necessarily free" to be his own tyrant, inexorably imposing upon himself the peculiar combination of insights associated with his peculiar combination of experiences."[86] The other impulse is to seek rhetorical consensus among "members of our species....through various *media* of symbolism."[87]

Physicalists seek rhetorically consensus through the common body of assumptions they share. Nelson Goodman articulates the monopolistic physicalist position as ambiently reductive:

> the physicalist who maintains that one system, physics, is preeminent and all-inclusive, such that every other version must eventually be reduced to it or rejected as false or meaningless. If all right versions could somehow be reduced to one and only one, that one might with some semblance of plausibility be regarded as the only truth about the only world.[88]

It must be noted however, that reductionism although used pejoratively at times, is inevitable in every form of human enterprise. As a quasi-rhetorician, this writer brings in his own terministic screens that simultaneously shed insights while at the same time blindside him from seeing other sides that other screens might yield. Scientists are not free from these screens either, however, it is instructive to note the presence of these screens. The quest for determinate knowledge yields perspectives that tend to provide an overarching view of the nature of things, by resorting to the kind of categorizations that Aristotle postulated through his systematic study of analytics.

Thomas Nagel notes in *The View from Nowhere*:

> Even if we acknowledge the existence of distinct and irreducible perspectives, the wish for a unified conception of the world doesn't go away. If we can't achieve it in a form that eliminates individual perspectives, we may inquire to what extent it can be achieved if we admit them. Persons and other conscious beings are part of the natural order, and their mental states are part of the way the world is in itself. From the perspective of one type of being, the subjective features of the mental states of a very different type of being are not accessible through subjective imagination or through the kind of objective representation that captures the physical world. The question is whether these gaps can be at least partially closed by another form of thought, which acknowledges perspectives different from one's own and conceives of them not by means of the imagination. A being of total imaginative flexibility could project himself directly into every possible subjective point of view, and would not need such an objective method to think about the full range of possible inner lives. But since we can't do that, a more detached form of access to other subjective forms would be useful.[89]

Nagel highlights the problem of achieving an overarching epistemology, and suggests the need to find means of inquiry albeit partial, yet intellectually honest enough to admit the finiteness of human inquiry. Yet, the rhetorical appeal of a "unified conception" will not fade away in scientific circles. In that light, Turing himself articulates a terministic screen of physicalism in which the physical system is marked by determinate limits: "the human computer is supposed to be following fixed rules; he has no authority to deviate from them in any detail."[90] The Turing test implies that both the human computer and the digital computer follow some particular neurological script although the pathways of a conversation seem random. Turing explains away the non-quantifiable aspects of mental states as irrelevant to the project of constructing intelligent machines. Such a position is consistent with the overall philosophy of eliminative materialism or physicalism.

Thus far, scientific ethos is established dialogically by validating the message through the credibility of the rhetor, and also through seeking validation from a larger matrix of scientific institutional practices – in this case, the tenets and practices of materialism. Robert K. Merton describes the ethos of science as:

> that affectively toned complex of norms which is held to be binding on the man of science. The norms are expressed in the forms of prescriptions, proscriptions, preferences and permissions. They are legitimized in terms of institutional values. These imperatives, transmitted by precept and example and reinforced by sanctions are in varying degrees internalized by the scientist, thus fashioning the scientific conscience....Although the ethos of science has not been codified, it can be inferred from the scientific moral consensus of scientists....in countless writings on the scientific spirit and moral indignation directed towards contravention of the ethos.[91]

Ethos thus becomes abstracted from the discourse of the rhetor who is certified as an expert, and also from the matrix of intellectual and social conditions that facilitate the emergence of a particular perspective. The Turing test and the assertions that Turing makes are rhetorically legitimated through the philosophy of physicalism or materialism blended with behaviorism.

Concluding Unscientific Postscript

John Milton in Book I of *Paradise Lost* declares his intentions for writing his neo-Homeric epic in the following fashion:

> Illumine, what is low raise and support;
> That to the heighth of this great argument
> I may assert Eternal Providence,
> And Justify the ways of God to men.[92]

In almost Miltonesque style, Turing attempts to justify the ways of an intelligent machine before humans.[93] Symbolic AI (Chapter 5) and Connectionism (Chapter 6) are in some fashion or the other worthy progeny of Turing's legacy.

Notes

[1] James Boswell, *Life of Johnson* (London: Oxford University Press, 1960), 90.
[2] Samuel Taylor Coleridge, *Specimens of the Table Talk of Samuel Taylor Coleridge* (London: Oxford University Press, 1851), 187.
[3] Thomas Carlyle, *On Heroes and Hero-Worship, & the Heroic in History* (Berkeley: University of California Press, 1993), 40.
[4] Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago: The University of Chicago Press, 1996).
[5] Turing also was one of the key figures who helped unravel the key to the German U-boat Enigma cipher that played an influential role in helping the Allies beat the Axis powers in WWII.
[6] Andrew Hodges, *Alan Turing: The Enigma* (New York: Walker & Company, 2000), 530.
[7] A. M. Turing, "Computing Machinery and Intelligence," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachussetts: A Bradford Book, MIT Press, 1997), 29-56.
[8] Turing, "Computing Machinery and Intelligence," 38.
[9] Edwin Tenney Brewster, *Natural Wonders Every Child Should Know* (New York: Grosset & Dunlap, 1912), 239.
[10] Hodges, *Alan Turing*, 63.
[11] Hodges, *Alan Turing*, 63-64
[12] Jack Emery, *Breaking the Code* (London: BBC productions, 1997).
[13] Brian Greene, *The Elegant Universe: SuperStrings, Hidden Dimensions and the Quest for the Ultimate Theory* (New York: Vintage Books, 2000), 341.
[14] Greene, *The Elegant Universe*, 341
[15] Hodges, *Alan Turing*, 65 (look for Eddington original quote)
[16] Hodges, *Alan Turing*, 66 (look for Eddington original quote)
[17] cited in Hodges, *Alan Turing*, 66.
[18] William Graham, *The Creed of Science* (London: Kegan Paul, 1881), 122.
[19] J. Loeb, *Comparative Physiology of the Brain and Comparative Psychology* (New York: Putnam, 1900), 214.
[20] J. Loeb, "The Significance of Tropisms for Psychology," in *The Mechanistic Conception of Life*, ed. D. Fleming (Cambridge, Massachusetts: The Belknap Press of Harvard University, 1964), 40-41.
[21] Stuart G. Shanker, "Descartes' legacy: the mechanist/vitalist debates," in *Philosophy of Science, Logic and Mathematics in the 20th century*, vol ix (London: Routledge, 1996), 348.
[22] cited in Donald Michie, "Turing's test and Conscious Thought," 35.
[23] Andy Clark, *Being There: Putting Brain, Body and World Together Again,"* (Cambridge, Massachusetts: MIT Press, 1999), 83.
[24] Clark, *Being There*, 83.
[25] A, Damasio and H. Damasio. "Cortical systems for retrieval of concrete knowledge: The convergence zone framework," in *Large Scale Neuronal theories of the Brain*, ed. C. Koch and J. Davis (Cambridge, Massachusetts: MIT Press, 1994), 138.
[26] Clark, *Being There*, 139.
[27] Stuart G. Shanker, "Introduction," in *Philosophy of Science, Logic and Mathematics in the 20th century*, Routledge History of Philosophy, Volume ix, ed. Stuart G. Shanker (Routledge: London, 1996).

[28] Bertrand Russell, "On Scientific Method in Philosophy," in *Mysticism and Logic* (New York: Norton & company, 1929).
[29] Bertrand Russell, *The Problems of Philosophy* (Oxford: Oxford University Press, 1959), find page number.
[30] Shanker, "Introduction," 4
[31] Ludwig Wittgenstein, *Culture and Value*, trans. P. Winch (Oxford: Basil Blackwell, 1980), 15.
[32] Ludwig Wittgenstein, *Wittgenstein Lectures: Cambridge 1930-1932*, Ed., D. Lee (Oxford: Blackwell, 1980), 26.
[33] Ludwig Wittgenstein, *Philosophical Investigations*, trans. G. E. M. Anscombe, (New York: Macmillan Press, 1968), 47. Russell derided Wittgenstein's project as one that would reduce philosophy "at best, a slight help to lexicographers, and at worst, an idle tea-table amusement." (Bertrand Russell, *My Philosophical Development*, London: George Allen and Unwin 1959, p. 161).
[34] F. E. Browder, *Mathematical Developments Arising from Hilbert Problems, Proceedings of Symposia in Pure Mathematics*, vol. 28 (Providence: American Mathematical Society, 1976), 1-34.
[35] cited in Shanker, *Philosophy of Science, Logic and Mathematics*, 17-18.
[36] cited in J. van Heijenoort, *From Frege to Gödel* (Cambridge, Massachussets: Harvard University Press, 1967), 152-182
[37] Dale Cyphert, "Strategic Use of the Unsayable: Paradox as Argument in Gödel's Theorem." *Quarterly Journal of Speech* 84(1998): 80-93.
[38] Kurt Gödel, "On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems I" in *Frege and Gödel: Two Fundamental Texts in Mathematical Logic,"* trans and ed., Jean Van Heijenoort (Cambridge: Harvard University Press, 1970), 83-108.
[39] Shanker, *Philosophy of Science, Mathematics and Logic*, 27.
[40] G. T. Kneebone, *Mathematical Logic and the Foundations of Mathematics* (New York: Dover, 2001), 358.
[41] Hodges, *Alan Turing*, 101.
[42] Alan Turing, "On Computable numbers, with an application to the *Entscheidungsproblem*," *Proceedings London Mathematical Society*, 43, 544-546.
[43] Turing, "On Computable numbers," 544-546.
[44] Donald Michie, "Turing's test and Conscious Thought," in *Machines and Thought: The Legacy of Alan Turing*, vol. 1, ed. Peter Millican and Andy Clark (Oxford: Oxford University Press, 1996), 28.
[45] A. M. Turing, "Lecture to the London Mathematical Society, 20 February, 1947" in *A.M.Turing's Ace Report of 1946 and other papers,"* Vol.10, Charles Babbage Institute Reprint Series for the History of Computing (Cambridge, Massachusetts: MIT Press, 1986), 106-124.
[46] A. M. Turing, "Lecture to London Mathematical Society," 122-123.
[47] Turing, "Lecture to London Mathematical Society," 123.
[48] Turing, "Lecture to London Mathematical Society," 123.
[49] Turing, *Lecture to the London Mathematical Society*, 124.
[50] J. Von Neumann, "General and Logical theory of Automata," in *Cerebral Mechanisms in Behavior-the Hixon Symposium* (New York: John Wiley, 1948), 303.
[51] Turing, "Computing Machinery and Intelligence," 32.
[51] Turing, "Computing Machinery and Intelligence," 33.
[52] Turing, "Computing Machinery and Intelligence," 33.
[53] Turing, "Computing Machinery and Intelligence," 33.
[54] Turing, "Computing Machinery and Intelligence," 33.
[55] Turing, "Computing Machinery and Intelligence," 34.
[56] Turing, "Computing Machinery and Intelligence," 37.
[57] Ajit Narayanan, "Intentional Stance and Imitation Game," in *The Legacy of Alan Turing: Machines and Thought, Vol. I*, ed. Peter Millican and Andy Clark (Oxford: Oxford University Press, 1999), 63-80 (76-77).
[58] David Marr, *Vision* (San Francisco: W. H. Freeman, 1982).

[59] Kenneth Burke, *Language as Symbolic Action: Essays on Life, Literature and Method* (Berkeley: The University of California Press, 1966).
[60] Burke, *Language as Symbolic Action*, 45.
[61] Burke, *Language as Symbolic Action*, 45.
[62] Turing, "Computing Intelligence and Machinery," 38.
[63] Turing, "Computing Intelligence and Machinery," 29.
[64] Daniel Dennett, 'Consciousness,' in *The Oxford Companion to the Mind*, ed. R. L. Gregory (Oxford: Oxford University Press, 1987), 162.
[65] Turing, "Computing Intelligence and Machinery," 31.
[66] Turing, "Computing Intelligence and Machinery," 39.
[67] Turing, "Computing Intelligence and Machinery," 39.
[68] J. R. Lucas, "Minds, Machines, and Gödel: A Retrospect," in *Machines and Thought: The Legacy of Alan Turing*, vol. 1, ed. Peter Millican and Andy Clark (Oxford: Oxford University Press, 1996), 103-124 (119).
[69] Turing, *Lecture to London Mathematical Society*, 122-123.
[70] Turing, "Computing Machinery and Intelligence," 53.
[71] Aaron Sloman, "Beyond Turing Equivalence" in *Machines and Thought: The Legacy of Alan Turing*, vol. 1, ed. Peter Millican and Andy Clark (Oxford: Oxford University Press, 1996), 179-219 (181).
[72] Sloman, "Beyond Turing Equivalence," 185-186.
[73] Alan Gross, *The Rhetoric of Science* (Cambridge, Massachusetts: Harvard University Press, 1996), 52.
[74] Turing, "Computing Intelligence and Machinery," 29-56.
[75] Turing, "Computing Intelligence and Machinery," 39.
[76] Turing, "Computing Intelligence and Machinery," 40.
[77] Turing, "Computing Intelligence and Machinery," 40.
[78] Turing, "Computing Intelligence and Machinery," 41.
[79] Turing, "Computing Intelligence and Machinery," 42.
[80] Turing, "Computing Intelligence and Machinery," 42.
[81] Aristotle, *On Rhetoric: A Theory of Civic Discourse*, trans. George A. Kennedy (New York: Oxford University Press, 1991).
[82] Aristotle, *On Rhetoric*, 1.2.4.
83 Helen Constantinides, "The Duality of Scientific Ethos: Deep and Surface Structure," *Quarterly Journal of Speech*, 87(2001): 61-72.
84 Willard Van Orman Quine, *Word & Object* (Cambridge, Massachusetts: MIT Press, 1960), 4.
85 Ian Hacking, *Representing and Intervening* (Cambridge: Cambridge University Press, 1983).
[86] Burke, *Language as Symbolic Action*, 52.
[87] Burke, *Language as Symbolic Action*, 52.
[88] Nelson Goodman, *Ways of Worldmaking* (Indianapolis: Hackett Publishing Company, 1978), 4-5.
[89] Thomas Nagel, *The View from Nowhere* (New York: Oxford University Press, 1986), 17.
[90] Turing, "Computing Intelligence and Machinery," 32.
[91] Robert K. Merton, "The Normative Structure of Science," *The Sociology of Science: Theoretical and Empirical Investigations* (Chicago, Illinois: The University of Chicago Press, 1973), 267-278.
[92] John Milton, *John Milton: The Complete Poems*, ed. John Leonard (London: Penguin Books, 1998), 121.
[93] John R. Searle critiques the Turing test in his piece "Minds, Brains and Programs" in Haugeland, *Mind Design II*. This writer will cover Searle's arguments in Chapter 5.

CHAPTER V

THE MIND IS A MACHINE

> If in the perception of corporeal phenomena external stimuli strike and impinge on the instruments of the senses, and corporeal passivity precedes mental activity – a passivity which stimulates mental activity and calls up the dormant forms in the mind – if, I say, in perceiving corporeal phenomena the mind is not passively affected, *but judges of its own power the experience subjected* (emphasis mine) to the body, consider the case of beings which in their mode of perception are free from all corporeal influence. They can rouse their mind to activity without having to react to external stimuli in order to perceive things. By this argument, therefore, a multiplicity of kinds of knowledge has been given to different substances. Mere sensation without any kind of knowing has been given to animals that have no power of movement, like mussels and other shellfish which grow on rocks. Imagination has been given to animals which do have the power of movement and which appear to have some will to choose or avoid things. *Reason belongs to the human race* (emphasis mine), just as intelligence belongs only to divinity –
>
> Boethius[1]

> Today, our machines are still simple creations, requiring the parental care and hovering attention of any newborn, hardly worthy of the word "intelligent." But within the next century they will mature into entities as complex as ourselves, and eventually into something transcending everything we know – in whom we can take pride when they refer to themselves as our descendants... *We are very near to the time when virtually no essential human function, physical or mental, will lack an artificial counterpart* (emphasis mine). The embodiment of this convergence of cultural developments will be the intelligent robot, *a machine that can think and act as a human* (emphasis mine) – Hans Moravec, a roboticist.[2]

Plato, Boethius and Descartes who respectively represent the intellectual zeitgeists of three major ages in history share one thing in common; namely a strong commitment to rationalism as a preeminent epistemology. The ascendancy of reason as a mode of knowing over emotivist or aesthetic accounts of the world paved the way for the development of modern science. Plato, Boethius and Descartes undeniably enamored by the powers of reason, subscribed to the view that the mind or intellect was an autonomous and superior entity, distinct from the corporeal and sensory body, and that rationality was unique to humans. The resurgence of scientific materialism eclipsed the substance-dualist rationalism that valorized the dichotomy between the mind and the body. Materialism, despite its diverging trajectories, is unified in its insistence that the mind *qua* mind is a product of mechanistic, neurological activity in the physical brain. In effect, the physicalist accounts of mind espoused by thinkers as diverse as Turing or Moravec revolve around a common theme – namely, that mental activity is not some mysterious, insuperable phenomenon but instead a tangible phenomenon with a material basis.

Artificial Intelligence, thus, became the proud progeny of scientific materialism. In other words, if one could create the material conditions in machines to display intelligent behavior – machines can engage in behavior that would require intelligence if performed by humans. As Marvin Minsky remarks, one can "build a mind from many little parts, each mindless by itself. Each mental agent by itself can only do some simple thing that needs no mind or thought at all. Yet when we join these agents in societies…this leads to true intelligence."[3] Intelligent behavior is construed as a manifestation of the interaction among material constituents, which in turn serve as the building blocks for mental activities. Such a rhetorical and philosophical move from mentalism to materialism facilitated the birth of Artificial Intelligence.

Early models of Artificial Intelligence were heavily influenced by symbolic or mathematical logic. The sentential epistemology of Symbolic logic entails a view that thought can be represented through formal propositions. Natural languages can be translated into discrete symbols and into formal propositions. James Crosswhite suggests the rhetorical dimensions of formal logic are as follows:

> It does this partly to disambiguate natural language, partly to remove anything but its formal features. By doing this, logic gains univocity, simplicity, and the power that comes from treating propositions in purely formal fashion. People can very quickly reach agreement about the formal relations among propositions, much less quickly about the strength of natural language arguments. The universal audience of logic is a reflection of this fact.[4]

The development of formal languages with disambiguated referents, under early models of AI, reinforced the view floated by early thinkers that thought can be represented through a set of discrete symbols. Symbolic AI postulates a symmetrical relationship between thought and the symbol by means of which it is symbolized. Just as thought is considered to be a by-product of mental activity in a material, neurological structure, semantics is considered to be a by-product of a syntactical structure. The underlying assumption being, if the syntax is sound, semantics will take care of itself.

Symbolic AI entails a top-down imposition of a rules-based, prescriptive grammar of intelligent behavior. Symbolic AI is based on formal languages, formal logic and construes the human activity of intelligence as a product of symbol manipulation. Formal languages follow a normative and denotative referential theory of meaning. Such an approach has obvious advantages, in that intelligence is formalizable and hence replicable and verifiable. Furthermore, a top-down conception of normative language would imply that the structure of a sentence regulates its accompanying meaning to circumvent the ambiguities inherent in the use of natural languages. Yet, its strength also happens to be its weakness. The counter argument to the 'pro' position may run as follows: at least in the realm of natural languages, the polysemic view of language implies that semantics is a dynamic field such that the shifting and evolving meanings of words could entail alterations in the structure as well. To put it simply, the tail could wag the dog as well. (For instance, the denotative meaning of the word 'dog' differs from the multiple connotations the same word assumes. For instance, the sense in which the term 'dog' is used in the statement "you are the man now, dog" is quite different from the statement "that freaky guy is following me around like a dog"). Such ambiguities would imply that the process of description would be open-ended and not foreclosed as formal languages would seem to suggest. Crosswhite explicates on the limitations of formal language as follows:

> It is oblivious to the ambiguities and resonances and moods of natural language since they are lost in the translation into a formal language. This (universal audience of logic) audience also lacks any sense of the importance or meaning of what is being argued. In some situations, real people will not reach agreement about how to translate between natural and formal languages. Some paragon audiences will refuse altogether to argue about some issues. Again, the point is that logic's universal audience is somehow *wrong* or misconceived. The point is that from some angles it is a very particular audience, not a universal one. It embodies many of our notions of reasonableness, if fails to embody others.[5]

Yet, the precise grammar of symbolic language has been key to developing the idea of symbol-manipulation in computers. At the heart of symbolic AI is the notion that intelligence is largely a product of symbol-manipulation. Bearing this in mind, the purpose of the chapter is to explore the rhetorical conception of thought in symbolic AI via interrelated concepts: a) formalization, b) mathematization and c) the mechanization of mind and discuss how the mechanistic metaphor (mind is a machine) plays out as a terministic screen.

The Discourse of Symbolic AI

Symbolic Artificial Intelligence is based on the assumption that intelligence is a function of symbol manipulation. From this perspective, the human brain can be seen as a symbol system. John Haugeland suggests that symbolic AI is "predicated on the idea that systems can be built to solve problems by reasoning or thinking them through in this way, and, moreover, this is how people solve problems."[6] According to symbolic AI, the mind "just *is* a computer with certain special characteristics – namely, one with internal states and processes that can be regarded as explicit *thinking* or *reasoning*."[7] The idea that the mind just is a computer or at least functions like one in articulated by Turing in his manifesto for machine intelligence. Turing's universal machine essentially is based on the conception that the ranges of activities that are computable by humans or any other machine are coextensive with the capabilities of the Turing machine. In his lecture to the London Mathematical Society in 1947 Turing describes the breadth and scope of his universal machine as follows:

> Let us now return to the analogy of the theoretical computing machines with an infinite tape. It can be shown that a single special machine of that type can be made to do the work of all. It could be in fact be made to work as a model of any other machine. The special machine may be called the universal machine; it works in the following quite simple manner. When we have decided what machine we wish to imitate we punch a description of it on the tape of the universal machine. This description explains what the machine would do in every configuration in which it might find itself. The universal machine has only to keep looking at this description in order to find out what it should do at each stage.[8]

Turing showed how using formalized procedures, one could describe an operation in terms of Boolean symbols 0 and 1. If the machine is provided with instruction tables as to how to interpret the information, computers can simulate intelligent behavior: "Actually one could communicate with these machines in any language provided it was an exact language, i.e. in principle one should be able to communicate in any symbolic logic, provided that the machine were given instruction tables which would enable it to interpret that logical system."[9] Turing explicitly states that the "language in which one communicates with these machines, i.e. the language of instruction tables, forms a sort of symbolic logic."[10]

Herbert Simon drawing on Turing's insights of instruction tables developed the notion of symbol structures:

> The primitives of mind, at the level I wish to consider, are symbols, complex structures of symbols, and processes that operate on symbols. The simplest among these processes require tens to hundreds of milliseconds for their execution. Simple recognition of a familiar object takes at least 500 milliseconds. At this level, the same software can be implemented with radically different kinds of hardware – protoplasm and silicon among them....at this level of aggregation conventional computers can be, and have been, programmed to represent symbol structures in a manner that parallels, step by step, the way the human brain does it. The principal evidence for my thesis are programs that do just that. These programs demonstrably think.[11]

Human thought processes effected through symbol manipulation are considered analogous to computers. The isomorphism is explicated in such a way to preempt criticism that the analogy between machines and humans might be inexact. Human thought or behavioral capabilities are to a certain extent conditioned by interactions in the brain, and thus the symbolists hypothesize that humans are programmed (invoking the metaphor of mechanism). In Simon's own words: "I cannot forbear...a brief comment on one of the commonest objections (to the notion that computers can think): that computers do only what they are programmed to do. The conclusion would only follow if it were true that human beings, when they are thinking, do not do what they are programmed to."[12] This argument of Simon is really not strong here especially when applied to humans – for instance, a person may suffer from a case of sexual addiction (be it neurological/hormonal) but still has the ability not to act out on these urges which simply means that humans can depart from scripts ("our nature") be they neurological, hormonal or whatever the case may be.

The physical symbol system hypothesis postulated by Allen Newell and Herbert A. Simon is at the heart of the symbolic AI project: "A physical symbol system has the necessary and sufficient means for general intelligent action."[13] Physical symbol systems are construed as systems that "clearly obey the laws of physics – they are realizable by engineered systems made of engineering components," and are "not restricted to human symbol systems."[14] Two notions central to symbol systems are a) designation and b) interpretation – designation as the rubric implies pertains to an expression that designates an object, "if given the expression, the system can either affect the object itself or behave in ways depending on the object," and interpretation suggests that the system can interpret the expression if the expression designates a process that can be carried out.[15]

Newell and Simon assert that symbols lie at the root of intelligent action, and that human brains by themselves are physical symbol systems. The authors postulate the hypothesis as being in accordance with the laws of qualitative structure: "All sciences *characterize the essential nature of the system* (emphasis mine) they study. These characterizations are invariably qualitative in nature, for they set the terms within which more detailed knowledge can be developed."[16] To justify their extrapolation, the authors refer to four exemplar models in the natural sciences that have yielded effective results with a high level of generalizability, namely the cell doctrine in biology (cells are the building blocks of all living organisms), plate tectonics (the earth's surface constitutes a collection of huge plates, the misalignment of any of these plates can cause cataclysmic events), the germ theory of disease (Pasteur suggested that most diseases are caused by minute unicellular living organisms) and the doctrine of atomism. All these four models shed a lot of insight into the essential nature of the systems they represent. The authors extrapolate this analogy to say that thought is a derivative of symbol-manipulation, therefore if one were to look for intelligent action one should closely study the symbol system. Symbol systems have the necessary and sufficient means for producing intelligent action:

> By "necessary" we mean that any system that exhibits general intelligence will prove upon analysis to be a physical symbol system. By "sufficient" we mean that any physical symbol system of sufficient size can be organized further to exhibit general intelligence. By "general intelligent action" we wish to indicate the same scope of intelligence as we see in human action: that in any real situation, behavior appropriate to the ends of the system and adaptive to the demands of the environment can occur, within some limits of speed and complexity.[17]

The symbolic paradigm operates under the assumption that thought can be formalized by symbols, and these symbols are in turn presented as the language of thought. Underlying the symbolic paradigm is the view that there are laws of thought that govern the language of thought. These rules are essentially "symbolic structures" that are supported by a physical symbol system (a computing device for manipulating symbols), in turn supported by lower implementation levels in a computing device. Paul Smolensky explains the underlying rationale behind the symbolist paradigm as follows:

> The idea is that, eventually, if we were to get low enough down in a physical symbol system, we would see something like neurons. In other words...we just have to figure out how to relate neural structures to the low implementation levels of a physical symbol system, and then we understand the relation between neural structures and mental structures.[18]

Through a combination of empirical science and analogical reasoning, symbolists establish an isomorphism between the human brain and the digital computer. However, it must be mentioned that most symbolists have however focused on higher-order intellectual activities like mathematics and so forth – and are more concerned with 'intelligence' per se and establishing isomorphism without necessarily paying attention to architectural differences. Eventually, somehow the differences will be bridged with the bigger and faster machines; or that is what we are told. Newell and Simon hypothesized that although the human brain and the digital computer are different in structure and mechanism, they can be seen as two different instantiations of a single species of device and thus share a common functional description, at the level of symbol manipulation.[19]

In general, the computational model of mind under symbolic AI rests on at least three interrelated and inseparable principles: a) the formalization of thought, b) the mathematization of thought and c) the mechanization of mind.

Formalization is contingent on the idea of a formal system. The notion of a formal system is largely borrowed from arithmetic and algebra. Solving arithmetic or algebraic problems involve manipulation of tokens according to definite rules not unlike games in a closed system where there are a finite number of calculable positions. The tokens usually stand for something, and hence are referred to symbols. To decipher what a body of symbols mean is not unlike the process of finding rationality in a body of behavior, whereby one looks for consistent and reliable pattern that renders the act of interpretation possible. Thus formalization is similar to an act of translation that entails two types of interpretation: a) intentional and b) semantic. Intentional interpretation pertains to the orienting of particular operations to achieve a particular desirable goal or behavior in a consistent and intelligible manner. Semantic interpretation, given the situation, pertains to the attribution of meaning to a body of symbols so that what they mean turn out to be consistently reasonable. Philip E. Agre suggests that when thought is formalized at least two patterns emerge:

> A word that once referred to something in the world now refers to a structure in the computer...A word that once referred to an activity conducted by agents in the world now refers to a process occurring entirely in the computer. Examples include "search," all verbs for operations on data structures...and many predicates on the internal operations of technical entities.[20]

Agre further elaborates on the mentalistic formalizations of symbolic AI as follows, "if agents need to think about the world, put analogs of the world in the head.

If agents need to act in situations, put data structures called "situations" in the head. If agents need to figure out what might happen, put simulations of the world in the head."[21] The formalization of thought occurs by breaking an idea into its primitive parts, and formalizing those primitive parts by means of logical operations that represent these parts. Simon posits the view that

> The successive levels in the architecture of nature are not arbitrary. Most complex systems are hierarchical and *nearly decomposable.* Consider a building divided into rooms, which are, in turn, divided into cubicles...In a hierarchical system of this kind, we do not have to consider the behavior at all levels simultaneously. We can model the cubicles, the rooms and buildings semi-independently...the behavior of nearly decomposable systems can be analyzed mathematically.[22]

Simon extrapolates the 'building' metaphor while speaking about the human mind: "because mind has shown itself to behave as a nearly decomposable system, we can model thinking at the symbolic level...without concern for details of implementation at the 'hardware' level, whether the hardware be brain or computer."[23]

Formalization entails at least two features; a) breaking something down into its constituent parts and b) establishing formal relations between the constituent parts, such that the interaction between the parts would suffice for a general explanation about the whole. If one were analyzing a particular behavior, one would start with a description of that behavior by breaking it down into the several facets that compose the behavior and then axiomatizing the behavior as a representative sample of the physical system it is representing. Frank Rosenblatt explains and critiques the symbolist impulse as follows:

> The implicit assumption (of the symbol manipulating research program) is that it is relatively easy to specify the behavior that we want the system to perform,

> and that the challenge is then to design a device or mechanism which will effectively carry out this behavior...It is both easier and more profitable to axiomatize the *physical system* and then investigate this system analytically to determine its behavior, than to axiomatize the *behavior* and then design a physical system by techniques of logical synthesis.[24]

In other words, instead of asking the question what kind of logical structure should a system have in order to exhibit some particular property, one should ask what can kind of system can facilitate the emergence or evolution of that property. (Thus, the boundary lines between the programmers (Symbolic AI) and the networkers (connectionists) were drawn).

Formalization can also be seen in the transformation of natural languages into a string of symbols whose meanings operate in a closely bound, finite problem space. From a rhetorical perspective, formal languages display the propensity to operate with both the "conduit metaphor" and "metonymy" as discursive orientations. At this juncture, a digression is necessary in order to preempt criticism pertaining to the apparent conflation between a metaphor (considered figurative) and an actual science (considered literal). I would like to state that metaphors could be seen as topographic models of discourse. In other words, a metaphor can be construed as a generative tool of discourse whose scope extends way beyond the literal/figurative dichotomy. Paul N. Edwards conception of discourse and its subsequent influence in shaping reality is particularly instructive:

> Discourse goes beyond speech acts to refer to entire field of *signifying* or *meaningful practices*: those social interactions – material, institutional, and linguistic – through which reality is interpreted and constructed for us and with which human knowledge is produced and reproduced. A discourse, then, is a way of knowledge, a background of assumptions and agreements about how reality is to be interpreted and expressed, supported by paradigmatic

> metaphors, techniques, and technologies and potentially embodied in social institutions.[25]

More specifically, the 'mind is a machine' emerges as a paradigmatic or generative metaphor in symbolic AI.

The generative metaphor vis-à-vis symbolic AI is particularly a conduit metaphor. George Lakoff and Mark Johnson explain the underlying assumptions behind a conduit metaphor as follows:

> the linguistic expressions are containers for meanings aspect of conduit metaphor entails that words and sentences have meanings in themselves, independent of people and contexts. The part of the metaphor that says linguistic expressions are containers for meanings entails that words (and sentences) have meanings, again independent of contexts and speakers. These metaphors are appropriate in many situations – those where context differences don't matter and where all the participants in the conversation understand the sentences in the same way. These two entailments are exemplified by sentences like *The meaning is right there in the words.*[26]

In other words, the meanings of terms are considered isolatable and separable from the dynamic contexts of everyday life and are posited as being "contained" within the words. The separation of language from its dynamic contexts of everyday speech is presumably made possible through the rules of grammar. Following the lead from Noam Chomsky and others, symbolists began speaking of a "grammar" in any language as comprising a finite set of rules which decides what strings or words may constitute acceptable sentences.[27] A computational grammar entails a systemic, epistemological orientation towards the world whereby semantics is preserved through syntax – and to a large extent, meaning is purportedly frozen unencumbered by the vagaries of shifting patterns of signification that are characteristic of most natural languages. Formalization is largely based on mathematical linguistics whereby words and strings of words are "detached from their meaning," such that mathematical linguistics becomes the study of "*meaningless* words and sentences."[28]

As stated earlier, the symbols in the computer are construed to represent context-free independent objective features and thereby eliminate ambiguity and foreground precision. One would veer away from making descriptions such as the car travels fast; instead one would specify the exact speed at which a car travels, thereby making a statement free of subjective interpretation. Hubert Dreyfus (symbolic AI's bête noire) assesses the precision-laden, disambiguous nature of symbolic manipulation as follows (although he does not use the term 'container' metaphor per se):

> The precision essential to a computer's way of manipulating symbols constitute both a great advantage and a severe limitation. Since what the symbols in a computer represent must be absolutely precise, and the programmer must be absolutely clear as to what he lets each symbol mean, the attempt to write a computer program inevitably exposes hand-waving, fuzzy thinking, and implicit appeals to what everyone takes for granted. Submitting to this rigor is an immensely valuable discipline. The analytic power of a computer used as a logic machine also has its limitations, however....The computer, if used to simulate logical thinking, can only make inferences from lists of facts. It's as if, in order to read a newspaper, you had to spell out each word, finds its meaning in the dictionary, and diagram every sentence, labeling all parts of the speech. Brains don't seem to decompose either language or images this way, but logic machines have no choice.[29]

What is representable, what is simulatable takes precedence over how humans interact in actual, real-world settings where communication practices are not always scripted. The obvious advantage is that the imperfection of human imprecision fades away under the rigorous formalism of symbol-arrangement and symbol-manipulation. Agre elucidates the rhetorical omissions of symbolic AI as follows, "the object of inquiry was not the individual in society or the person in the world, but the self-sufficient inner realm of the mind. The conceptual autonomy and infinite generative power of mental computations has played the same role in the computational theory of mind that the transcendence of the soul played for so long in philosophy."[30] Just as mental states within a discrete, well-defined computational system are considered self-sustaining and self-perpetuating, meaning is considered 'contained' within a self-sustaining and self-perpetuating structure.

Metonymy as a topographic orientation also prefigures in the strategic discourse of formalization prevalent in symbolic AI. Lakoff and Johnson explain metonymy as "a referential function, that is, it allows one entity to *stand for* another...For example, in the case of the metonymy THE PART FOR THE WHOLE there are many parts that stand for the whole. Which part we pick out determines which aspect of the whole we are focusing on."[31] Metonymic concepts are systematic in that they help rhetors organize their thoughts and ideas around salient, representative aspects. Metonymies are great tools of practical reasoning anchored in an anthropomorphism of some sort or the other. Examples of anthropomorphisms are replete in everyday discourse: phrases such as 'get your *behind* over here' (part for the whole), 'Remember the *Alamo*' (place for the event), 'she bought a Martin's' (producer for product), 'the motorcycle is ready to go for a spin' (object used for the user), 'the *Army* wants to reinstitute the draft' (institution for people responsible) are just a few examples for metonymies.

In AI literature, accomplishing a particular task-domain in a problem-space becomes a metonymy for mental activity, mental activity for intelligence, intelligence for the mind, and the mind for a person. Metonymies can be seen as formalized levels of implementations in a hierarchy of representation. Turing's postulation of a serial computer has at least three main parts: a) store, b) executive unit and c) control. The Storage Unit is a metonymy for memory (where 'instruction tables' are located) as a whole, the Executive Unit is a metonymy for the various operations involved in a calculation, and Control (the duty of the control is to see that the instructions provided by the "instruction tables" in the store are followed) as a metonymy for the information processing mechanism in the brain. Metonymies are heuristic and hermeneutic tools that come in handy for modeling. Agre elucidates on the nature of these representations in the following fashion:

> Scientific inquiries based on technical modeling should be guided by a proper understanding of the nature of models. A model is, before anything else, an *interpretation* of the phenomenon it represents. Between the model and the putative reality is a research community engaged in a certain discursive operation, namely, glossing some concrete circumstances in a vocabulary that can be assimilated to certain bits of mathematics. The discourses within which this process takes place are not transparent pictures of reality; nor are they simply approximations of reality. On the contrary, *such discourses have elaborate structures and are thoroughly metaphorical in nature* (emphasis mine). These discourses are not simply ways of speaking; they also help organize mediated ways of *seeing*. They provide the vocabulary for formulating models, interpreting results, and then choosing among revised models.[32]

The metonymic path that symbolists undertake is a journey of representation, whereby the process of symbol manipulation in computation is synonymous or interchangeable with the mental operations of the actual mind: "at the symbolic level the human mind is fundamentally a serial machine, accomplishing its work through temporal sequences of processes, each typically requiring hundreds of milliseconds for execution."[33]

As stated earlier, the impulse of symbolic AI from a metonymic perspective is to conflate symbol-manipulation for mental activity, mental activity for intelligence, intelligence for the mind, and the mind for a person.

From a rhetorical standpoint, the mathematization of thought implies at least three things: a) the mathematization of physical reality (the external world is finite and calculable, thanks to Newton and other mechanists) b) the representation of thought through symbols (the birth of mathematical logic) c) the conflation of thought with computation and symbol-manipulation (Turing's equivalence between the human computer and the digital computer).

Lord Kelvin in his lecture on units of electrical measurement, enunciates the importance of "numerical reckoning" in gaining epistemological certainty in the physical sciences:

> In physical science a first essential step in the direction of lengthening any subject, is to find principles of numerical reckoning, and methods of practically measuring, some quality connected with it. I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of *science*, whatever the matter may be.[34]

In effect, Lord Kelvin's now immortalized comments finely summarizes not only the impulse of the scientific spirit, but encapsulates the centrality of mathematical thought in the sciences. Although Lord Kelvin among other distinguished scientists recognized the probabilistic nature of physical reality, yet it was precisely the language of mathematics that lent credence to their scientific descriptions of reality. Lord Kelvin's predecessor Joseph Fourier asserted the cardinality of mathematical analysis in enhancing understanding of the world as follows:

> Mathematical analysis is as extensive as nature itself; it defines all perceptible relations, measures times, spaces, forces, temperatures; this difficult science is formed slowly, but it preserves every principle which it has once acquired; it grows and strengthens itself incessantly in the midst of many variations and errors of the human mind. Its chief attribute is clearness; it has no marks to express confused notions. It brings together phenomena the most diverse, and discovers the hidden analogies which unite them...It makes them present and measurable, and seems to be a faculty of the human mind destined to supplement the shortness of life and the imperfection of the senses; and what is still more remarkable, it follows the same course in the study of all phenomena; it interprets them by the same language, as if to attest the unity and simplicity of the plan of the universe, and to make still more evident that unchangeable order which presides over all natural causes.[35]

In other words, a universe of elegant and calculable simplicity necessitates a consistent and rigorous methodology to uncover its physical laws and processes. Mathematics indeed was put forward as the answer. In the fact, the self-enclosed and self-regulated world of Cartesian *mathesis* found its entelechy in the extension of mathematics to other domains of human inquiry. The rhetorical appeal of mathematics lay largely in its self-referential coherence and consistency that in turn provided a vertistic model for other physical sciences to emulate. Such an assertion is not to gloss over the differences apparent between the disciplines, but the insuperable sense of verifiability and accuracy that mathematics exudes has been hard to resist. In Albert Einstein's words, the ascendancy of Newtonian classical mechanics can be traced to the primacy accorded to mathematics, with the differential equation projected as "the natural expression of the elementary in physics:"

> It is to be observed that this theoretical system (Newtonian mechanics) is by nature atomistic and mechanical; all activity was to be thought of as purely mechanical – that is, simply as the movement of material particles according to Newton's laws of motion...In order to give his system mathematical form, Newton had to discover the concept of the differential coefficient, and to enunciate the Laws of Motion in the form of differential equations – perhaps the greatest intellectual stride that it has been ever granted for any man to make. For this purpose partial differential equations were not necessary, and Newton made no methodical use of them. But they were necessary for the formulation of the mechanics of deformable bodies...Thus the partial differential equation came to theoretical physics *as a servant, but by degrees it became its master* (emphasis mine). This process began in the nineteenth century, when, under pressure of facts of observation, the undulatory theory of light gained acceptance. Light in empty space was conceived to be a vibration of the ether as itself as a conglomeration of material particles. Here for the first time the partial differential equation appeared as *the natural expression of the elementary in physics.*[36]

On one hand, there have been contextual and paradigmatic shifts in apprehending the physical world – yet methodologically speaking mathematics has retained pride of place as the queen of the sciences because it provides useful tools for calculation and measurement.

Although there has been a paradigmatic shift from Newtonian classical mechanics to a quantum mechanical account of the universe, theoretical physics relies heavily on mathematics and experimental sciences for validation. Even the experimental sciences rely heavily on mathematic modeling to analyze the structure or microstructure of various phenomena under investigation. Thus instead of being a mere tool to unravel physical reality, mathematics became the preeminent way of knowing.

If the physical reality of the universe could be expressed mathematically, the same analogy of mathematical form could extend to the physical reality of the mind as well. Thereby, classical approaches in Artificial Intelligence can be seen as offshoots of developments in mathematics. Pioneering work in symbolic logic done by Boole, Frege, Russell among others culminated in the fruition of AI as a serious, scientific inquiry. Leibniz suggested that it was possible to represent 'common sense' knowledge with mathematical reasoning. Boole systematized the idea by developing a system of mathematical logic that enabled the representation of logical propositions with algebraic symbols.

If one were interested in looking at symbolic AI, from the perspective of intellectual history (or the history of ideas), situating George Boole in a rhetorical and historical context will be particularly helpful. During the nineteenth century, a rancorous debate took place between the metaphysician Sir William Hamilton and Augustus de Morgan, a symbolic logician. Hamilton, a neo-Aristotelian logician repudiated mathematical logic vis-à-vis the training of the moral mind – since he felt that mathematical logic would merely foster an uncritical acceptance of whatever premises were proposed and also generate a rather mechanistic, unreflective view of demonstration. Hamilton further suggested that logical metaphysics pertained to real existences and causes, while mathematics merely described and measured without an emphasis on first principles.[37] Boole taking sides with de Morgan countered Hamilton's claims by stating that all logical or scientific reasoning can be expressed more efficiently with symbols, elaborating that the symbolism provided by Aristotelian logic was a foreshadowing of the symbolism to come. In Boole's words,"according to my view of the nature of philosophy (metaphysics), *Logic forms no part of it*...we ought no longer to associate Logic and metaphysics, but Logic and Mathematics.'[38]

Boole's dissociation of logic from metaphysics entailed an emphatic first step towards the mathematization of logic, and subsequently the mathematization of thought. Boole essentially does three things: first, he asserts that signs and symbols are central to knowledge and thought; second, he considers all thought to be within the domain of logic, and third, logic belongs to the domain of mathematics or vice-versa. Therefore, thought falls within the domain of mathematics. In his words:

> Now the expression of Thought here implied is the office of signs or symbols of which the words of common speech are the most familiar examples. And of all the systems of signs this indeed is the most important...Number, magnitude and their relations, the so-called affections of space the ultimate forces and elements of the material universe so far as they are present known to us, have been represented and the thought of which they are the objects expressed by signs. In this its highest conception therefore Logic might be said to be the Philosophy of *all* thought which is expressible by signs whatever the object of that thought, whatever the nature of the those signs may be...There is a philosophy of signs which governs and explains all their particular uses and applications, - which is equally manifested in the forms of ordinary speech and in the symbolical language of mathematics. The perfect idea of Logic is not that of a mere system of rules but a philosophy from which as from a common stem all sciences whose method is deductive are developed and with which they all stand in vital connection.[39]

The signs were considered as keys to mental operations, such that Boole rested his treatise 'An Investigation of the Laws of Thought' on two fundamental presuppositions: a) the operations of the mind are subjected to general laws and b) these laws are "mathematical in their form," and that they are "actually developed in the essential laws of human language. Wherefore the laws of symbols of logic are deducible from a consideration of the operations of the mind in reasoning."[40] In effect, it is noteworthy that Boole begins his 'Investigation of the laws of thought,' with an explicit declaration of intent:

> to investigate the fundamental laws of those operations of mind by which reasoning is performed; to give expression to them in the symbolic language of a calculus and, upon this foundation, to establish the science of logic and construct its method; to make that method itself the basis of a general method for the application of the mathematical doctrine of probabilities; and finally, to collect from the various elements of truth brought to view in the course of these inquiries some probable intimations concerning the nature and constitution of the human mind.[41]

How did Boole go about investigating the nature of thought? Boole believed that there was a science behind thought processes not unlike the fact that there is a science behind the material world. However, a distinct demarcation is made between the general laws of nature and the laws of the mind, in that the laws of nature or our apprehension of physical reality is dependent on observations while "knowledge of the laws of mind does not require as its basis any extensive collection of observations."[42] Instead of relying on copious observations and on a chain of inferences to unravel the laws of the mind, Boole suggested that the science of logic, in itself, was sufficient to shed insight on understanding the nature of mental operations:

> But if the general truths of Logic are of such a nature that when presented to the mind they at once command assent, wherein lies the difficulty of constructing the Science of Logic? Not, it may be answered, in collecting the materials of knowledge, but in discriminating their nature, and determining their mutual place and relation. All sciences consist of general truths...Let us define as fundamental those laws and principles from which all other general truths of science may be deduced, and into which they may all be again resolved. Shall we then err in regarding that as the true science of Logic which, laying down elementary laws, confirmed by the very testimony of the mind, permit us then to deduce by uniform processes, the entire chain of its secondary consequences, and furnishes, for its practical applications, methods

> of perfect generality? Let it be considered whether in any science viewed either as a system of truth or as the foundation of a practical art, there can properly be any other test of the completeness and the fundamental character of its laws, than the completeness of its system of derived truths, and the generality of the methods which it serves to establish.[43]

More specifically, a science of logic could uncover the fundamental laws of reasoning in the symbolic language of calculus. Although Boole recognizes that laws of the symbols of logic and those of algebra are independent of each other, he suggests that "there is to a considerable extent an exact agreement in the laws by which the two classes of operations are conducted."[44] Employing analogical and practical reasoning, Boole establishes an isomorphism of equivalence between the two by "certain general principles founded in the very nature of language, by which the use of symbols, which are but elements of scientific language, is determined."[45]

Thus Boolean logic or Boolean algebra can be seen as an effort to systematize thought through the use of mathematical symbols. Boole elucidates that "the intellectual operations...conducted by means of their instrument, language, are formal, and are concerned not with the nature of the individual object of thought but only with the scientific notion under which that object is apprehended."[46] An elaborate system of scientific notations based on formal laws of representation is considered a more effective way of systematizing thought than by merely analyzing ordinary language. In Boole's words:

> The ground of every system of notation employed in reasoning is the formal character of reasoning itself. If the process of inference is independent of the particular meaning of the concepts involved and depends only upon the general notion which those concepts manifest and thereupon only as it furnishes the basis of intellectual operations and of formal laws it is at once suggested to us that *we express concepts not as in ordinary language by words* (emphasis

> mine) the special meaning of which may through association of ideas interfere with our perception and application of the purely formal laws to which they are subject but by symbols in the employment of which the formal law and not the special meaning is present to the mind.[47]

Boole implies that normal language interferes with our perception due to its inherent instabilities. A formal language with mathematical symbols operates in a hermetic space that is devoid of the vagaries of human perception. The normal descriptions of ordinary language, especially correct grammatical descriptions are based on the laws of grammar in their syntactical arrangement of words. Yet, the semantic aspects of language given the polysemic nature of language, shows that language does not exist in a closed system whereby the elements of signification are cut off after a certain point. It is precisely against the ambiguous nature of natural languages, that Boole proposes his mathematical model of language. The imperfect and confused shadows of natural languages fades away as the bright and perfect light of mathematical and algebraic logic dawns on human thought:

> The excellence of a notation consists in this that it expresses directly by the elementary symbols the elementary concepts operations and relations of the system of thought to which it belongs...The perfection of the language of Algebra is due to the circumstance that it has been found possible thus to determine beyond all question or contradiction the ultimate elements of thought in the system to which the language is applied.[48]

How is such perfection acquired, one might ask? To which Boole would respond that in the system of algebra there exists an elementary concept known as a number, to which one could apply four operations such as addition, subtraction, multiplication and division. The conception of the number can be represented by the Arabic numerals, while the operations of addition, subtraction, multiplication and division can be represented by the respective signs such as +, - and so forth. When one expresses these elements by signs, it follows that "all their combinations possible in thought will be expressible by combination of signs in subjection to formal laws which represent the laws of combinations of the original elements."[49]

For instance if the number 6 was multiplied by 8 the result is the same as the number 8 multiplied by 6, namely 48. If x and y represent any two numbers and if xy represents the product by multiplying the number x with the number y. Therefore, one could say that xy is equal to yx. If one were to represent these entities with the law of addition, one would say x plus y is equal to y plus x. If one were to employ the formal laws of operations in which addition and multiplication are mixed, one would get $x(y + z) = xy + xz$.[49] Boole proceeds to say that the elements of Logic "are not less definite than those of Algebra," and the elementary concept in Logic is that of the "class" – of which "four elementary operations by which concepts of class can be modified viz. Addition, Subtraction, Composition and Abstraction but one fundamental relation expressed by propositions viz. the relation of identity."[50] The rationale for appropriating algebraic symbols for expressing logical propositions is explained as follows:

> It is by no means essential to adopt in the expression of the forms of Logic the symbols employed in Algebra. It is however a matter of convenience to do so. For although the ideas embodied in the symbols of Algebra are for the most part different from those which the same symbols would embody if employed for the expression of propositions in Logic yet does there exist between the two sciences such a fundamental relation (however it my be explained) that the formal laws of the symbols are with one exception the same in the two systems.[51]

Having said so, Boole states one could denote classes of things with letters and the four elementary logical operations of addition, subtraction, composition and abstraction can be represented with the same signs used to denote the respective arithmetical operations of addition, subtraction, multiplication and division. Boole's significant contribution that was eventually extrapolated to the development of computer science are as follows:

> Let *x+y* denotes the class formed by adding the members of the class *x* to those of the class *y* supposed distinct. E.g. if *x* represent "Trees" and *y* "Herbs" let *x+y* represent "Trees and herbs." – Let *x* – *y* denote the class formed by subtracting from the class *x* the class *y* supposed to be wholly contained therein. E.g. if *x* represent "Men" and *y "Byzantinians," let x – y represent "men who are not Byzantinians."* Let *x* x *y* or *xy* denote the class whose members are common to the classes *x* and *y*. Thus if *x* denote "Flowers" and *y* "White things" let *xy* denote "White flowers." – Let *x/y* represent that class of things from which if we select those which belong to the class *y* we shall obtain the class *x*.[52]

W. V. Quine remarks that the Boole's postulation of the class was quite influential in broadening the scope for logicians and mathematicians in studying "all subject matter."[53] The ingenuity of Boole's contributions rest largely in the fact that he "allows contributions which have no direct logical interpretation" and applies them to discrete classes of objects.[54]

Boole also introduces the binaries 1 and 0, with 1 representing the universe of conceivable objects and 0 representing the null class. Boole elucidates his rationale as follows:

> The ground of this selection is that the symbols 0 and 1 are subjected to the same formal laws when thus interpreted in Logic as when employed in Arithmetic. Hence the expression 1 – *x* will denote that entire class of things which remains after taking away from the universe the class denoted by *x* i.e. it will denote the class of things which are not members of the class *x*. – Let the

> placing of any expression within brackets or under a vinculum denote that it is to be treated according to the same laws as if it were a single letter representing a class. Thus let $y(1-x)$...denote the class of things which consists of all individuals that are found in the class y but not in the class x.[55]

By categorizing objects into distinct classes, Boole demonstrates the possibility of representing propositions from natural language by means of mathematical symbols. Such a representation not only signifies the ability to mathematize thought, but also carries with the efficacy of mathematical precision that is lacking in natural languages. An example that Boole provides is the proposition: Stars are celestial bodies which either are self-luminous and do not shine by reflected light or shine by reflected light and are not self luminous- Stars by *s*, Celestial bodies by *c*, Self-luminous bodies by *a*, Bodies shining with reflected light by *r*. Therefore, the above-mentioned proposition can be expressed symbolically through the equation $s = c\ [\ a(1 - \mathrm{r}) + \mathrm{r}(1 - \mathrm{a})]$ where $\mathrm{a}(1 - \mathrm{r})$ stands for the class of things which contains all entities found in class *a* (self-luminous bodies) but not in class *r* (bodies that shine with reflected light) and where $\mathrm{r}(1 - \mathrm{a})$ stands for the class of things which contains all entities found in class *r* (bodies that shine with reflected light) but not in class *a* (self-luminous bodies).[56]

Although the example provided above is fairly simple, Boole incontrovertibly demonstrates that one could symbolize any logical proposition as long as the elements of the propositions falls within the purview of a "class." Furthermore, Boole suggests that one could construct an entire algebraic system just with the numbers 0 and 1:

> Now there exists two numbers viz. 0 and 1 which besides satisfying the general laws of Algebra satisfy also the above special formal laws of logic. If then we construct an Algebra in which the only particular symbols of number understood an Algebra in which the only particular symbols of number shall be 0 and 1 and in which every general symbol as *x*, *y* etc. shall be understood to

> admit only of the above special determination (i.e. it being given that x is a literal symbol belonging to the dual Algebra it shall thence be understood that x means either 0 or 1...) the formal laws of such an Algebra will be identical with those of Logic when expressed by symbols...And hence the processes of the dual Algebra...will be formally identical with the process of Logic expressed by symbols.[57]

Thereby, Boole sowed the seeds for the germination of binary arithmetic whose scientific character via the principle of symbolization led to the development of computer science. In effect, the scientific character of Boolean logic is derived from the application of mathematical symbols that are consistent and not subject to the semantic quirks of ordinary language. One of Boole's contemporaries Richard Whately in his influential *Elements of Logic* describes the close interconnection between logic and mathematics as follows:

> There is in fact a striking analogy in this respect between the two sciences. All Numbers (which are the subject of Arithmetic) must be numbers of *some things*, whether coins, persons, measures, or any thing else; but to introduce into the science any notice of the *things* respecting which calculations are made, would be evidently irrelevant, and would destroy its scientific character; we therefore proceed with arbitrary signs representing numbers in the abstract. So also does Logic pronounce on the validity of a regularly-constructed argument, equally well, though arbitrary *symbols* may have been substituted for the Terms; and consequently, without any regard to the things signified by those Terms. And the possibility of doing this...is a proof of the strictly scientific character of the system.[58]

It is essentially the scientific character of mathematical logic that paved the way for figures such as Alan Turing. Boolean logic and the subsequent development in symbolic logic became central to the development of digital logic circuits, largely based on binary arithmetic expressed with the numerals 1 and 0 and also to the "account of human reasoning as the calculation of truth values of logical propositions."[59] Although many current programmers and developers might be unaware of the extent to which symbolic logic helped influence the advent of computer science, it is only fair to say that historians of computer service would be doing a great disservice if they do not mention Boole's contributions. Biographer Desmond MacHale best expresses Boole's enduring legacy:

> It threw a great deal of light on the nature of mathematics; it opened up possibilities of an extension of the subject into totally new and unexpected areas – classical mathematics had concentrated on the notions of shape and number and even when symbols were employed, they were generally interpreted in terms of number. Boole had now introduced the notion of interpreting symbols as classes or sets of objects, a concept breathtaking in scope because it meant that the *study of all well-defined sets of objects now came under the realm of mathematics* (emphasis mine)...By enlarging the horizons of mathematics so enormously, Boole unwittingly (but perhaps subsconsciously, wittingly) highlighted a topic that has come to influence virtually every aspect of present-day life – the storage and processing of information, which in turn has led to the development of computer science. Not alone is Boole's algebra the 'correct' and most economical tool for handling information, but the electronic machines which now do the work actually operate according to principles determined by that self-same algebra. Boole has been called the 'Father of Symbolic Logic' and the 'Founder of Pure Mathematics,' but he is just as deserving of the title, 'Father of Computer Science.'[60]

Other symbolic logicians who deserve honorable mention include Frege and Russell. The Boolean project of mathematizing thought reaches its pinnacle with Turing. The mathematization of thought becomes computation, with Turing being the first mathematician and logician who organized the modern digital computer on paper. Turing considered human thought to be a product of symbol-manipulation. It is also safe to say that "thought" and "computation" could be used interchangeably if one were to use Turing's vocabulary. In his words:

> The behavior of the computer (human) at any moment is determined by the symbols which he is observing, and his "state of mind" at that moment. We may suppose that there is a bound *B* to the number of symbols or squares which

> the computer can observe at one moment. If he wishes to observe more, he must use successive observations. We will also suppose that the number of states of mind which need to be taken into account is finite.[61]

In effect, the isomorphism between human thought and computation via symbolic-manipulation is considered a given and also as a working hypothesis to generate models of intelligent behavior in non-human physical systems. Therefore, the distinction between artifacts and natural systems are construed as arbitrary:

> This distinction is, however, purely a social construct. If humans are viewed as physical systems, then their design and construction of artifacts is a physical process like any other, and the physical results of this process are as natural as the humans that constructed them. Computational behavior emerges when systems of either sort are measured in particular ways. The scientific study of the dynamic and algorithmic similarities in the behavior of humans and some of their constructions need not take this socially motivated distinction into account.[62]

Turing provides a model with which to generate symbol processing or manipulation through a digital computer, implying that mental activity is calculable and replicable in non-human physical systems as well:

> We may compare a man in the process of computing a real number to a machine which is only capable of a finite number of conditions...which will be called "*m*-configurations." The machine is supplied with a "tape" (the analog of paper) running through it, and divided into sections (called "squares") each capable of bearing a "symbol." At any moment there is just one square...bearing the symbol which is "in the machine." We may call this square the "scanned square." The symbol on the scanned square may be called the "scanned symbol." The "scanned symbol" is the only one of which the machine is so to speak "directly aware." However, by altering its *m*-configuration the machine can effectively remember some of the symbols which it has "seen" (scanned) previously. The possible behavior of the machine at any moment is determined by the *m*-configuration...and the scanned symbol...[63]

In other words, information can be represented by the symbols in particular squares; the scanning of these squares is a form of symbol processing or manipulation.

Turing further elaborates on the mechanics of symbol-manipulation, by establishing equivalence between the human and machine processes of computation, the only difference being what humans do on paper, computers do on a tape:

> If at each stage the motion of a machine is *completely* determined by the configuration, we shall call the machine an "automatic machine" (or *a*-machine)...If an *a*-machine prints two kinds of symbols, of which the first kind (called figures) consists entirely of 0 and 1 (the others being called symbols of the second kind), then the machine will be called a computing machine. If the machine is supplied with a blank tape and set in motion, starting from the correct initial *m*-configuration...the motion of the machine, the number of the scanned square, the complete sequence of all symbols on the tape, and the *m*-configuration will be said to describe the *complete configuration* at that stage.[64]

Symbol-manipulation is efficiently carried out when one introduces instruction tables for the machine to follow. Such a task entails mechanical reasoning that both machines and humans are adept at, by following instruction tables. Turing draws a comparison between the two as follows:

> We suppose...that the computation is carried out on a tape; but we avoid introducing the "state of mind" by considering a more physical and definite counterpart of it. It is always possible for the computer (human) to break off from his work, to go away and forget all about it, and later to come back and go on with it. If he does this he must leave a note of instructions (written in some standard form) explaining how the work is to be continued. This note is the counterpart of the "state of mind." We will suppose that the computer works in such a desultory manner that he never does more than one step at sitting. The note of instructions must enable him to carry out one step and write the next note. Thus the state of progress of the computation at any stage is completely determined by the note of instructions and symbols on the tape.[65]

Turing's description of computation vis-à-vis machines is architecturally elegant and has historically been largely beneficial in the development of logic circuits. Turing's account is also highly instructive in understanding the nature of mechanical reasoning, due to the systematic and step-by-step elaboration of how one could build such a machine. In effect, the Turing equivalence between mechanical and human reasoning has been at the heart of the symbolic, AI project. Yet, it might be helpful to call for a brief and temporary suspension of the conflation of distinctions - not to diminish or negate the AI project - but only to elicit more inquiry into the nature of human computation before equivalences are asserted.

Aaron Sloman's critique of Turing equivalence that runs is particularly insightful:

> The mathematical concept of 'computation' is the only well-defined concept of computation. It is concerned purely with formal structures. This point can be obscured by the process-product ambiguity. A process of computation may produce a *trace*, for example a long division presented on paper. Both the process and its enduring trace can be called computations, but in different senses. The formal concept of computation involves no notion of process, causation, or time, and is concerned only with the structural properties of such traces, no matter how they are produced. (Similar process-product ambiguities are associated with: 'proof,' 'derivation,' 'calculation,' 'analysis,' 'design,' 'construction').[66]

Computation is primarily a formal process based on the principle of symbolization, during the abstraction of which ideas such as causation, time among others are ignored. The process of computation in itself pays attention only to what is being represented on paper or the program, and not necessarily to how it is being represented. Sloman further elaborates:

> The formal notion of computation, equally applicable to physical processes and non-physical mathematical structures, does not on its own enable us to...explain human behavior. An abstract instance of computation (e.g. a huge Gödel number) cannot make anything happen. This shows that a computation in the formal sense is not a *sufficient* condition for being an intelligent *behaving* system, even though the formal theory provides a useful conceptual framework for categorizing some behaving systems. For instance, it establishes limits to what is possible and provides a framework for studying space-time complexity requirements and trade-offs. For the purpose of

> construction and explanation of intelligent systems, we need to combine computational ideas with the idea of a machine with causal powers.[67]

Furthermore, Turing's model of computation implies some sort of a Laplacian deterministic worldview where both minds and machines operate in a finite and closed system. A single statement that would characterize Turing's philosophy would read like this, namely, *thought equals computation*. The Turing equivalence, according to Sloman, breaks down when one takes into account the fact that physical systems are probabilistic and not deterministic:

> If, as physicists tell us a) no physical system has a totally determinate observable state, and b) transitions from one state to another are probabilistic rather than deterministic, then even physical objects in themselves, as opposed to our measurements of their behavior, cannot be treated as Turing equivalent computers, since (a) and (b) contradict requirements for Turing machines. Actual computers are built so as to minimize the large-scale effects of (a) and (b). Failure to do this completely leads to malfunctions, though mechanisms such as self-correcting memory devices reduce their impact.[68]

If thought and computation are the same, the world we live in is rather deterministic. If the world we live in is rather probabilistic, thought and computation are not always the same in that there is plenty of room for non-algorithmic elements as well.

The nature of the rift between the closed/determinist and open-ended/probabilistic camps is rather wide and needs to be taken into account, to have a proper understanding of arguments both for and against the mechanistic thesis. The mathematician von Neumann characterization of modern, mathematical logic is an exemplar for those who would argue for Turing's equivalence between thought and computation:

> Throughout all modern logic, the only thing that is important is whether a result can be achieved in a finite number of elementary steps or not. The size of the number of steps which are required, on the other hand, is hardly even a concern of formal logic. Any finite sequence of correct steps is, as a matter of principle, as good as any other...In dealing with automata, this statement must be significantly modified. In the case of an automaton the thing which matters is not only whether it can reach certain finite result in a finite number of steps at all but also how many such steps are needed.[69]

Quantum Physicist Roger Penrose in *The Emperor's New Mind* argues against the strong deterministic thesis associated with the conflation 'the mind is a machine,' and instead prefers a probabilistic, quantum mechanical world where there is room for non-algorithmic elements as well:

> According to strong determinism, it is not just a matter of the future being determined by the past; *the entire history of the universe is fixed,* according to some precise mathematical schema, *for all time.* Such a viewpoint might have some appeal if one is inclined to identify the Platonic world with the physical world in some way, since Plato's world is fixed once and for all, with no 'alternative possibilities' for the universe...As a variant of strong determinism, one might consider the *many-worlds* view of quantum mechanics. According to this, it would not be a *single* individual universe-history that would be fixed by a precise mathematical scheme, but the totality of myriads upon myriads of 'possible' universe-histories that would be so determined. Despite the unpleasant nature of such a scheme and the multitude of problems and inadequacies that it presents us with, it cannot be ruled out as a possibility.[70]

To be more precise, Penrose appears to be arguing against the 'thought is computation' thesis and states that one cannot have a comprehensive account of human minds without taking into account the role of consciousness:

> Science seems to have driven us to accept that we are all merely small parts of a world governed in full detail (even if perhaps ultimately just probabilistically) by very precise mathematical laws. Our brains themselves, which seem to control all our actions, are also ruled by these same precise laws. The picture has emerged that all this precise physical activity, is in effect, nothing more than the acting out of some vast (perhaps probabilistic computation) – and hence, our brains and our minds are to be understood solely in terms of such computations...*In my own arguments I have tried to support this view that*

> *there is something that is missing from any purely computational picture* (emphasis mine)...Consciousness seems to me to be such an important phenomenon that I simply cannot believe that it is something just 'accidentally' conjured up by a complicated computation. It is the phenomenon whereby the universe's existence is made known. One can argue that a universe governed by laws that do not allow consciousness is no universe at all. I would even say that all mathematical descriptions of a universe that has been given so far must fail this criterion. It is only the phenomenon of consciousness that can conjure a putative 'theoretical' universe into actual existence.[71]

However, for the most part, the types of arguments that Penrose, John R. Searle[72] and others make are considered by advocates of strong AI as some sort of pre-scientific, mystification of the mind (Searle came up with his 'Chinese Rooms' thought experiment to counter Turing's imitation game – the CR thought experiment simply states that a person who does not know a lick of Chinese could merely manipulate Chinese symbols, based on instructions received from instruction tables, and this act of manipulation in itself does not constitute having an understanding of Chinese – and therefore Searle makes the claim that the imitation game is not sufficient to advance claims about cognition or thinking systems). In other words, symbolic representation or manipulation by itself is not adequate to account for intelligence, common sense or cognition.

Unlike Searle, Penrose and other detractors, the early AI community suggests that symbolic representation in itself is sufficient for generating intelligent behavior, and that the programs facilitating intelligent behavior can be ascribed with mental qualia. John McCarthy (remembered as the developer of Lisp) states that representation is an adequate index of intelligent behavior as long as the style of representation is adequate to generate intelligent behavior – McCarthy in an influential paper entitled 'Programs with Common Sense' theorizes about the ability to construct programs that are capable of manipulating in a "suitable formal language common instrumental programs:"

> The *advice-taker* is a proposed program for solving problems by manipulating sentences in a formal language...the main advantages we expect the *advice taker* to have is that its behavior will be improvable merely by making statements, telling it about its symbolic environment and what is wanted from it. To make these statements will require little if any knowledge of the program or the previous knowledge of the *advice taker*. One will be able to assume that the *advice taker* will have available to it a fairly wide class of immediate logical consequences of anything it is told and its previous knowledge. This property is expected to have much in common with what makes us describe certain humans as having *common sense*. We shall therefore say that *a program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows.*[73]

In other words by means of formal deductions from a given set of premises, the computer is said to have 'common sense' knowledge about the world, given the information that it is passed on to it by means of logical propositions.

Logical propositions enunciated in the computer programs are construed as tools to represent 'common sense' knowledge. John McCarthy and P. J. Hayes talk about the scope and possibility of such knowledge representation through a *Reasoning Program (RP)*:

> RP interacts with the world through input and output devices some of which may be sensory and motor organs (for example, television cameras, microphones, artificial arms) and others of which are communication devices (for example, teletypes or keyboard-display consoles). Internally, RP may represent information in a variety of ways. For example, pictures may be represented as dot arrays or as a list of regions and edges with classifications and adjacency relations. Scenes may be represented as lists of bodies with positions, shapes, and rates of motion. Situations may be represented by symbolic expressions with allowed rules of transformation."[74]

And with regard to representing other types of knowledge like sentences, McCarthy and Hayes state the following:

> All other data structures have linguistic descriptions that give the relations between the structures and what they tell about the world.
> The subroutines have linguistic descriptions that tell what they do, either internally manipulating data or externally manipulating the world.
> The rules that express RP's beliefs about how the world behaves...are expressed linguistically.
> RP's goals, as given by the experimenter, its devised subgoals, and its opinion on its state of progress are all linguistically expressed.
> We shall say that RP's information is adequate to solve a problem if it is a logical consequence of all these sentences that a certain strategy of action will solve it.[75]

Such linguistic data structures are considered epistemologically adequate representations of the real world, as long as the program is able to make formal deduction from a set of clearly spelt out premises and recursive procedures and thereby arrive at the desirable conclusion. McCarthy further elucidates that one could ascribe mental qualities to these 'common sense' programs:

> To ascribe certain *beliefs, knowledge, free will, intentions, consciousness, abilities,* or *wants* to a machine or computer program is legitimate when such an ascription expresses the same information about the machine that it expresses about a person. It is useful when the ascription helps us understand the structure of a machine, its past or future behavior, or how to repair or improve it. It is perhaps never logically required even for humans, but expressing reasonably briefly what is actually known about the state of a machine in a particular situation may require ascribing mental qualities or qualities isomorphic to them.[76]

McCarthy considers mathematical logic (although he calls for more inquiry in nonmonotonic reasoning, so that common sense programs can jump to conclusions even on grounds of insufficient evidence, as is common in natural languages) to be an adequate vehicle of epistemologically adequate representations of thought, such that the symbolic means of enunciating this representation via the computer program should be ascribed with mental qualities similar to the way in which humans are ascribed with mental qualities while representing the world internally or externally. In his words, all kinds of mathematical logic rest on two ideas - First, it must be "mathematically definite what strings of symbols are considered formulas of the logic," and second it must be "mathematically definite what inferences of new formulas from old ones are allowed. These ideas permit the writing of computer programs that decide what combination of symbols are sentences and what inferences are allowed in a particular language."[77]

Computer programs rely on mathematical logic at some level or the other, in that propositions from natural languages are symbolically represented in such a manner that the representations by themselves are considered vehicles of 'knowledge' about a specific domain. McCarthy goes on to add:

> Machines as simple as thermostats can be said to have beliefs, and having beliefs seems to be a characteristic of most machines capable of problem-solving performance...Mental qualities peculiar to human-like motivational structures, such as love and hate, will not be required for intelligent behavior, but we could probably program computers to exhibit them if we wanted to, because our common sense notions about them translate into certain programs and data structures. Still other mental qualities, e.g, humor and appreciation of beauty, seem much harder to model.[78]

McCarthy's acknowledgment of the difficulties in programming subjective qualia is consistent with the criticism directed against symbolic AI.

Mechanical reasoning is more readily formalizable, representable and simulatable. Given this assumption, it is also safe to say that machines are better at mechanical reasoning, unless there are technical malfunctions. Simon's Logic

Theorist is a good program to illustrate how computers are really efficient in solving theorems. In effect, one can even conjecture to say that machines are more efficient in spatial and mathematical reasoning and at tasks that value consistency and accuracy.

(A prime example of, may I say 'machine superiority' is the recent Chess match between Kasparov and Deep Blue. Game 6 is particularly noteworthy:

White – Deep Blue Black – Kasparov

1 e4 c6 2. d4 d5 3. Nc3 de 4. N:e4 Nd7 5. Ng5 Ngf6 6. Bd3 e6 7. N1f3 h6 8. N: e6 Qe7 9. 0-0 fe 10. Bg6+ Kd8 11 Bf4 b5 12. a4 Bb7 13. Re1 Nd5 14. Bg3 Kc8 15. ab cb 16. Qd3 Bc6 17. Bf5 ef. 18. R:e7 B:e7 19. c4 1: 0.[79]

Kasparov and Deep Blue were tied at 2.5 – 2.5 before game 6. Kasparov did not start the game strongly (Kasparov should have played Bd6 instead of h6 in move 7, thus not allowing his pieces to develop properly). In move 17, Deep Blue moves its Bishop to f5, offering a bishop gambit. What follows next made chess players all around the world drop their jaws in sheer astonishment – Garry Kasparov (the world's highest rated chess player and probably the greatest chess player as well, if not only next to Bobby Fischer) makes a move that rank amateurs would dread to commit. He fell for the gambit, by taking the bishop with his pawn leaving his queen defenseless. In move 18, Deep Blue takes Kasparov's queen leaving Kasparov no choice but to resign. Deep Blue wins the rematch 3.5 – 2.5.

The point of bringing in this game is a fine exemplar of human fallibility compared to the invulnerability of machines (except for instances of technical malfunctions). The type of mistake Kasparov made only goes to show that humans are far clumsier. It is unthinkable for Deep Blue to make the type of error that Kasparov committed in move 17. The counter-argument to this claim I made might be to say that one could program a computer to make mistakes. My response is why? It seems as if the programmer is placing the machine at a distinct disadvantage by doing so.

Yet, it must be said that there are many competent human Grand Masters who can still beat Deep Blue. However, it is safe to say that even the best human chess players in the future are going to be no match to chess playing computers since the processing power of computers are going to increase exponentially, making it very difficult for human chess players to compete at a competent level. In effect, Grand Master Viswanathan Anand best illustrates the contrasting styles between humans and machines in his assessment of the Kasparov-Deep Blue match:

> I eagerly waited to see the Kasparov vs. Deep Blue rematch. Deep Blue was stronger. Deeper to be precise. From my own experience, practical play exposes all sorts of weaknesses and strengths in my play that are hidden during preparation. Equally, the team behind Deep Blue must have benefited immensely from studying the six games played against Kasparov in 1996. And it would be faster. I can't tell the difference between 100 zillion positions and 497 zillion positions, but if it helped Deep Blue play stronger, so be it...Kasparov himself must have studied the game last year. However, humans can't change their style drastically like computers. On top of that, all his games were accessible to the Deep Blue team, while he was in the dark about Deep Blue. He had two options: to play like Kasparov or to play like "Mr. Anti Deep Blue." *The former runs the risk of playing to the strengths of the machine, the latter ends up as disoriented as the machine. Humans, too, play weaker in unfamiliar situations and though they may find their way around better, machines can compensate for that with brute force.*[80]

In spite of the contrasting styles, it is fair to state that the advent of Chess-playing software has greatly enhanced human performance, therefore the human versus machine dichotomy is not always helpful).

To state that computers are more adept at mechanical reasoning is not to preclude computers from performing well in other domains. However, the sentential epistemology of symbolic AI is narrowly domain-specific. The limitations of symbolic AI will be discussed at greater length in the next chapter (since connectionism claims to redress the omissions of symbolic AI) and also in the section entitled 'rhetoric of symbolic AI.'

Even while acknowledging the tremendous value in symbolic AI, one should not ignore the significant rhetorical omissions, namely the foreclosure of the process of signification. If one were to operate under the assumption that all thought is equivalent to computation, then one would essentially be saying that all the 'qualitative' aspects of human experience comes under the realm of computation. Such a position is consistent among proponents of the strong AI thesis who assert that even these so-called 'qualitative' aspects are only computational, namely that pains, joys, feelings of beauty, despair among many others are algorithmic, and hence simulatable. At the time of this writing, it seems to me that proponents are overpredicting the computational model even while reducing human experience into neat, algorithmic computations. Such a reduction would ignore the fact that humans live in a dynamic and rather open-ended world, filled with ambiguities and uncertainties, the way in which we navigate the problems of every day life do not follow any prescribed, calculable and inflexible predetermined code. Embracing the idea of human complexity does not necessarily undermine the notion that biological conditioning plays a role in human behavior, but only reinforces the view that narratives of reduction are also narratives of selection and omission.

The idea that human emotions can be programmed entertained by John McCarthy among others is certainly an enterprising and laudable one, yet it might be helpful to make meaningful distinctions before erasing them – to say that a simulation of thinking is also thinking (ontologically speaking) is plausible especially when one takes into account the nature of mechanical reasoning, but yet I am not certain yet that the simulation of pain is equivalent to pain (ontologically speaking) *in all cases*. At this point, it seems to me that there are instances where the computational model break down. Hypothetically speaking, let's assume that there is a computer that can explain and simulate the actual physical pain experienced by a human. Let's assume that the human is suffering from a terminal illness. Let's assume that the human has only a few more days to live, since this virus is in its advanced state destroying all his T-cells. The intelligent computer is very well able to explain what is going on with the human patient. Now take this same computer and give it the physical appearance of a humanoid robot and teach it to simulate the physical manifestation of the actual pain that the human patient is suffering – vividly simulating his facial expressions, the watery secretions of his lachrymal glands, his verbalization of frustration at the rejection and bigoted statements of some of his fellow homo sapiens. The question is whether the simulation of the pain is an actual expression of pain? If this question is not clear, one could rephrase the question in such a manner – is the simulated suffering of the AIDS patient by the humanoid equivalent to the real suffering experienced by the human? At this juncture, given my limited knowledge the rationalist in me would say 'no,' although the evocative appeal of such a virtual act is undeniable such that it is tempting to conflate the simulation with the real in a postmodern, Baudrillardian sort of fashion. And indeed, if it is real then one should wonder if it is ethical to inflict unjustifiable suffering on computers. There are myriads of other examples where the analogy of the simulation being equivalent to the 'real' breaks down. On the other hand, if a computer is programmed to love as in Spielberg's movie A.I, and interacts in a dynamic environment with other humans and other robots – it is so much easier to say that the simulation of love is equivalent to real love since love is actually an action word and not merely a propositional attitude. Programming computers to love and perform compassionate acts in real-time environments are much more rhetorically compelling. Under such circumstances even hard-nosed skeptics are going to significantly soften their calloused hearts.

What the future holds for computation is not my place to judge and I may very well be completely wrong in being overcautious. Yet at this stage it might be helpful to make non-dogmatic meaningful distinctions without necessarily blindly capitulating all human experience to the brute power of automated reasoning.

From a rhetorical perspective, a non-integrated approach of symbolic modeling of intelligence relies largely on a string of dissociations and abstractions; namely separating information from meaning, semantics from syntax, structure from action, sign from signification, system from process and lastly, agent from context. The theory of communication presupposed by this model is largely 'psychologistic' construed in behavioristic input-output, stimuli-response models. As Keith J. Holyoak and Paul Thagard illustrate:

> The web of culture that holds people together in social groups is construed from shared beliefs and feelings, knowledge of a common history, and a sense of place in the natural and social world. These strands provide the connections by which members of a society can communicate with one another. Myth and magic, rites and ceremonies, poetry and everyday conversation all form part of the web. A culture is built and maintained in large by symbolic stories and rituals, in which objects and events are given meanings that in various ways go beyond themselves. Analogy plays a prominent role in providing these extended meanings and thus in building and maintaining the web of culture.[83]

An over reliance on mathematical reasoning even while carrying with it the import of scientific rigor elides the complexities and instabilities inherent in natural languages, by valorizing a particular referent and by the same token suppressing the dynamics of the system of natural languages that enables the enunciation of the referent. On one hand, geometric reasoning or mathesis accords certainty, but yet, on the other hand it renders the world opaque and artificial by not embracing the instabilities of everyday life. Kenneth J. Knoespel articulates the importance of natural languages in the social construction of technology, by critiquing the impulse to suppress the model that natural languages provide as follows:

> The reification of geometry in architecture and technology has enormous implications for language. Once geometry becomes manifest in artifacts, these artifacts retain an authority radically different from that accessible to natural language. By virtue of their being manifested as physical objects they acquire what appears an autonomy separated from natural language. The apparent separation of both architecture and technology from language has great significance, for it works to repress the linguistic framework that has allowed them to come into being.[84]

To summarize, technology is a social construct just as much society is a technological construct in that technology has altered our perceptions of how we see the world and how we see the world. The advent of modern computers coincided with computational accounts of the mind, and by the same token, the social and cultural matrices surrounding World War II, the emergence of global capitalism, the information society among others engenders the possibility of such constructions. To ignore either one, is only at the detriment of the other. Thus, there is a need for more holistic approaches to questions such as intelligence and consciousness, which means that modelers who strive for verisimilitude, at whatever level they possibly can, should not only study mathematical logic but also pay attention to the sociological and cultural forces that make communication possible. More importantly, modelers should not only be good programmers but astute observers of human communication as well. However, I am not so sure if one can really capture all the vagaries and polysemic aspects of natural language.

The mathematization of thought is dependent on the mechanization of the mind, in that the mind is a material entity with a material basis and not a distinct metaphysical entity in the sense in which Descartes conceived the mind. The mechanization is based chiefly on a materialist thesis. In this section, I will survey a few of the historical arguments advanced in favor of materialism, and the subsequent mechanization of mind that can be seen as a rhetorical foreshadowing for the ‘mind is a machine’ computational metaphor. Materialism certainly has distinct trajectories, however all materialist accounts of mind are unified in their rejection of the distinctive Cartesian cogito.

Descartes, himself, a mechanist exempts the mind from the purview of mechanistic explanations. *Meditations II* enunciates what has become a manifesto for substance dualism:

> And first of all, because I know that all things which I apprehend clearly and distinctly can be created by God as I apprehend them, it suffices that I am able to apprehend one thing apart from another clearly and distinctly in order to be certain that the one is different from the other, since they may be made to exist in separation at least by the omnipotence of God; and it does not signify by what power this separation is made in order to compel me to judge them to be different: and, therefore, just because I know certainly that I exist, and that meanwhile I do not remark that any other necessarily pertains to my nature or essence, excepting that I am a thinking thing, I rightly conclude that my essence consists solely in the fact that I am a thinking thing (or a substance whose whole essence or nature is to think). And although possibly (or rather certainly, as I say in a moment) I possess a body with which I am intimately conjoined, yet because, on the one side, I have a clear and distinct idea of myself inasmuch as I am only a thinking and unextended thing, and as, on the other, I possess a distinct idea of body inasmuch as it is only an extended and unthinking thing, it is certain that this I (that is to say, my soul by which I am what I am), is entirely and absolutely distinct from my body, and can exist without it.[85]

The metaphysical Cartesian *cogito* is the exact antithesis of the physicalist/materialist accounts of mind; the latter accounts gained ascendancy due to the difficulty Cartesian metaphysicians encountered in explaining mind-body interaction. At another level, the Cartesian depiction of mind failed to account how mental processes originated. Implicit in Cartesian substance dualism is the assumption that thought *qua* thought is also a product of mental causation. Ensuing materialists considered such a mentalistic thesis circular and unsatisfactory. Although Descartes himself is a scientific rationalist and has been rather influential in ushering in a scientific/rational worldview, his consideration of *res cogitans* as extensionless places his account of mind in the realm of metaphysics.

The materialist Thomas Hobbes on the other hand, considered the body to be analogous to machinery of which thought was only a mechanistic manifestation. Hobbes' philosophy of mechanism suggests a human-machine isomorphism such that what is true for humans (namely that humans who are pieces of biological machinery have the capacity to think) is true for machines as well:

> Nature, the art by which God hath made and governs the world, is by the *art* of man, as in many other things, so in this also imitated, that it can make an artificial animal. For seeing life is but a motion of limbs, the beginning of which is in some principle part within; may we not say, that all *automata* (engines that move themselves by springs and wheels as doth a watch) have an artificial life? For what is the heart, but a spring; and the nerves, but so many strings and the joints but so many wheels, giving motion to the whole body, such as was intended by the artificer?[86]

Hobbes' belief in the unity of science led him to postulate the interaction between mind and body in physicalist terms, such that the mind cannot be excluded from the domain of science. A. P. Martinich explains the physicalist thesis as follows: "the only real things are bodies in motion...if one began with full knowledge of the smallest bodies and the laws of nature, then one could presumably deduce the whole course of the world, including the actions of human beings."[87] Hobbesian psychology entails an explanation of mental processes in mechanistic terms:

> By *Consequence*, or TRAIN of thoughts, I understand that succession of one thought to another, which is called (to distinguish it from discourse in words) *mental discourse*. When a man thinketh on any thing whatsoever, his next thought after, is not altogether so casual as it seems to be. Not every thought to every thought succeeds indifferently. But as we have no imagination, whereof we have not formerly had sense, in whole, or in parts; so we have transition from one imagination to another, whereof we never had the like before in our senses. The reason whereof is this. *All fancies are motions within us, relics of those made in the sense: and these motions that immediately succeed one another in the sense, continue also together after sense: insomuch as the former coming again to take place, and be predominant, the latter followeth, by*

> *coherence of the matter moved...the discourse of the mind, when it is governed by design, is nothing but seeking, or the faculty of invention...a hunting out of the causes, of some effect, present or past; or of the effects, of some present or past cause* (emphasis mine).[88]

In other words, just as the mechanics of motion can be explained through formal, causal explanation the dynamics of thought can be explained with such machine-like efficacy.

Another significant attack on the independence of the Cartesian cogito worth noting is from Gottfried Wilhelm Leibniz:

> The reason that Descartes alleged to prove the independence of our free actions by a pretended vivid internal feeling has no force. We cannot, strictly feel our independence, and we are not always aware of the causes, often imperceptible, on which our resolution depends. It is as if the magnetic needle found pleasure in turning towards the north; for it would believe it turned independently of any other cause, not being aware of the insensible movements of magnetic matter.[89]

Leibniz found an essential inconsistency in Descartes' reasoning. Descartes had commented about the nature of mind and body as follows:

> And this is the best way to understand the nature of the mind, and the distinction between it and the body. For *examining what we are* who suppose everything which is different from us to be false, we *clearly see* that no extension, nor figure, nor local motion, nor anything similar that is attributed to body, belongs to our nature, but only thought, which therefore we know prior to and more certainly than any corporeal thing; for this we already perceive, but so far we doubt the others.[90]

Leibniz attacks this argument as follows:

> It does not follow: I can assume or imagine that no bodies exist; but I cannot imagine that I do not exist, or that I do not think; therefore I am not a body, nor is thought an attribute of body. And I am astonished that such a distinguished man could have given so much credit to such a trivial sophism.[91]

Descartes' appeal to a 'clear and distinct' perception of the *Cogito* and his subsequent dissociation of the *cogito* from the body was based on an intuitionist thesis and not any hard evidence. Margaret Dauler Wilson suggests that Leibniz pictures Descartes reasoning inconsistently and incorrectly along these lines: "I know that I exist as a thinking thing; I do not know (i.e. can doubt) that matter exists; therefore the self that I know to exist is not material."[92]

However, Leibniz and Descartes shared common ground with regard to their belief in the immateriality of the soul whose existence was undiminished by the decaying of the corporeal body. Yet, Leibniz was a mechanist in that he posited the presence of a natural force that acted on corporeal bodies implying that there was a physical explanation for even seemingly incorporeal phenomena such as minds, reinforcing the view that there is indeed an undeniable union of the soul and the body:

> For why could not God give to a substance at the outset a nature or internal force which could produce in it in an orderly way...everything that is going to happen to it, that is to say, all the appearances or expressions it is going to have, and all without the help of any created thing? This is more likely since the nature of a substance necessarily requires and essentially involves some progress or change, without which it would have no force to act. *And as the nature of the soul is to represent the universe in a very exact way* (though with more or less distinctness), *the succession of representations which the soul produces for itself will naturally correspond to the succession of changes in the universe itself; just as on the other hand the body has also been adapted to the soul for the occasions when we think of the soul as acting externally* (emphasis mine)...Thus as soon as we see that this *Theory of Agreements* is possible, we see also that it is the most reasonable, and that it gives a wonderful sense of the harmony of the universe and the perfection of the works of God.[93]

Although Leibniz seeks ultimate validity in metaphysical first principles such as God, his belief in the unity and consistency of physical systems rendered a mechanistic conception of universe coherent. Leibniz's postulation of a monad or a simple substance that are corporeal yet animated 'intelligences' lends itself to a mechanistic explanation of the universe, namely that complex systems are built from simpler, self-regulating parts (Leibniz states 90 theses, only a few of which will be restated here):

> The *monad*...is nothing but a simple substance, which enters into composites; *simple*, meaning without parts.
> And there must be simple substances, because there are composites; for the composite is nothing but a collection, or *aggregatum*, of simples.
> Now, in that which has no parts, neither extension, nor shape, nor divisibility is possible. And so monads are the true atoms of nature; in a word, the elements of things.
> There is also no way in which it could make sense for a monad to be altered or changed internally by any other created thing. Because there is nothing to rearrange within a monad, and there is no conceivable internal motion in it which could be excited, directed, increased, or diminished, in the way that it can in a composite, where there is change among parts...
> It follows from what we have just said that natural changes in a Monad come from an *internal principle*, since no external causes could ever have an influence into its interior.
> Everyone must admit that perception, and everything that depends on it, is *inexplicable by mechanical principles*, by shapes and motions, that is. Imagine there were a machine by its structure produced thought, feeling, and perception; we can imagine it as being enlarged while maintaining the same relative proportions, to the point where we could go inside it, as we would go into a mill. But if that were so, when we went in we would find nothing but pieces which push one against another, and never anything to account for a perception. Therefore, we must look for it in the simple substance, and not in the composite, or in a machine. And that is all we can find within a simple substance, namely perceptions and their changes; and that is all that the *internal actions* of simple substances can consist in.
> Thus every organic body of a living being is a kind of divine machine or natural automaton.[94]

At this juncture, I would like to note that the samples of Leibnizian theses I have cited are far from being a comprehensive account of Leibniz's polymath epistemology. However, with respect to the project at hand, what is most salient about Leibniz's monads are the following assumptions: a) physical reality is composed of simple, primitive elements, b) these elements are self-sustaining and self-regulating and c) they operate with the consistency and regularity of a machine that has distinct, yet interrelated parts. Even if the final cause rested in the realm of metaphysics (due to Leibniz's deep religious sentiments) it is safe to infer that Leibniz's picture of the physical world is one of a gigantic machine, where the independent parts work together in a state of preestablished harmony.

On one hand, it is simplistic to speak of Hobbes and Leibniz in the same breath, largely due to their diverging metaphysical assumptions. Even so, both share the idea that 'matter can think' which happens to be true of French materialists such as Baron d'Holbach, Denis Diderot and Julien Offray de La Mettrie as well. Baron d' Holbach in *Systéme de la nature* pronounces the following:

> The system of spirituality, as it is understood today, owes all of its purported proofs to Descartes...He is the first one who established that whatever thinks must be distinguished from matter; from this he concludes that our soul or that which thinks in us is a mind, namely, a simple and indivisible substance. Wouldn't it have been more natural to conclude that since man, who is comprised of matter and whose ideas are only of matter, has the faculty of thinking, then matter can think?[95]

The self-sustaining nature of material systems that contained within themselves the seeds for intelligence made Cartesian accounts of mind unintelligible and unscientific. Denis Diderot similarly locates thought within material systems:

> Thinking is the result of sensibility, and according to me, sensibility is a universal property of matter, a property which is inert in elementary bodies, like motion in heavy bodies stopped by an obstacle, and a property which becomes active in the same bodies, due to their assimilation with a living animal substance.[96]

The term sensibility by itself is quite fuzzy, however it requires thinking or reflection which in itself is a product of matter.

La Mettrie tried to provide an account of materialism that is sufficiently rich enough to capture complexities associated with mental states by going into the structural details of the material brain, and in turn showing how the brain influences thought without seeking refuge in metaphysical idealizations:

> I can see only matter in the brain and only extension, as we have proved, in its sensitive part; when alive, healthy and well organized, this organ contains at the source of the nerves an active principle spread through the medullary substance. I can see this principle, which feels and thinks, being disturbed, falling asleep and dying with the body...If everything can be explained by what anatomy and physiology reveal to me in the medulla, what need do I have to forge an ideal being.[97]

Ann Thomson explains that "behind La Mettrie's claim that 'man is a machine' lies the fundamental affirmation that organized matter in motion produces intelligence under certain conditions, in animals as well as in humans."[98] In terms of structural organization, La Mettrie asserts that the terms 'humans' and 'machines' could be used interchangeably since humans are essentially systems of interweaving fibers and organs:

> But since all the soul's faculties depend so much on the specific organization of the brain and of the whole body that they are clearly nothing but that very organization, the machine is perfectly explained! For after all if even man alone has received the law of nature as his heritage, would he be any less a machine? Some wheels, a few springs more than in most perfect animals, the brain proportionately closer to the heart and thus receiving more blood, the same gift of reason or – how do I know? – unknown causes would always have produced that delicate conscience which is so easily wounded, that remorse which is no more foreign to matter than is thought, and in short all of the differences that are supposed here. So does the organization suffice to explain everything? Once again, yes. Since thought clearly develops with the organs, why should the matter which composes them not also be capable of remorse once it has acquired, with time, the faculty of feeling.[99]

La Mettrie in effect concludes his treatise with what was then considered a rather revolutionary statement: "Let us then conclude boldly that man is a machine and that there is in the whole universe only one diversely modified substance."[100]

Materialist thinkers such as the ones mentioned above prefigured the development of a fairly well-entrenched tradition of scientific thought whose suasory assertion lies in its rhetorical constitution of a measurable and calculable universe, whose operations invoke the metaphor of machinery, with its insistence on self-consistency, self-regulation and self-perpetuation. In this system, one should look for formal causal laws within the closely defined system, and not outside it – in that extraneous or nonscientific postulations were not necessary to explain processes that lend themselves to naturalistic explanations.

Newton's classical mechanics, Laplacian determinism, an evolving thread in scientific materialism, large-scale innovations in technology, the ushering in of the industrial revolution among other factors reinforced notions of a mechanistic universe, of which minds are material constituents: "from classical antiquity until relatively recently, the regularity of the universe was searched for and perceived in thematic harmonies. The idea that nature behaves systematically in the sense we understand it."[101]

From a rhetorically perspective, what is so compelling about the 'human is machine' metaphor? The clockwork precision of machinery, internal composition and consistency, the measurable interaction of interrelated parts, the integrity of an external structure, the principle of causation in that one could explain what causes a component to behave in a particular way, the notion of a closed system in a finite problem space among others imports a calculus of ratiocination that lends the language of scientific machinery more amenable to scientific theorizing than mere metaphysical pronouncements. If the mechanistic thesis could encompass all phenomena including mental processes, the mind as a machine metaphor can be seen as being apt and fitting.

On one hand, within a mechanistic paradigm one could find evidence to fit the model. Yet at another level, the model in itself is a rhetorical imposition, a lens with which materialist thinkers find particular affinity. Alan Gross suggests that the epistemological assumptions of science are rhetorically constituted through a "network of persuasive structures, patterns that extend upward through style and arrangement to invention itself, to science itself."[102] On a reflective note, it must be stated that much misunderstanding has been generated when one juxtaposes the term science with rhetoric; the common argument raised against this juxtaposition is that science has nothing to do with rhetoric. Interestingly enough, philosophers of science found themselves in the same boat not too long ago in that the practitioners; especially the hard-nosed experimenters asserted that they had little or no value for scientific philosophy (which is still unfortunately the case in considerable circles). It certainly did take a lot of persuasion to argue that a philosophy of science does indeed shed much light on the scientific method, in that a good practice entails a good grasp of some of the assumptions that go into the practice. By the same token, what can a rhetorical approach to science offer? It teaches us if nothing else, that the way in which rhetors use language plays a significant role in the way in which shape or fashion our perceptions of reality. Attempts to exclude rhetoric are only symptomatic of an ill-conceived distrust of the role of natural languages, even while detractors quite ironically use language and rhetorical strategies to discount the role of rhetoric. Taylor's account of the association between rhetoric and scientific thought is especially revealing:

> Because such a notion is likely to strike some as far too radical, I should be clear on what I take to be its most tantamount entailments...*To argue that science is constitutively rhetorical is not to deny the brute fact of empirical regularities in the natural world* (emphasis mine), for instance. I do not suggest that scientists make it up as they go. Certain recurrent patterns in the natural world, however (by themselves) do not constitute science (referring to scientific practice); they become science (scientific practice) only via the process of interpretation, hence reconstruction.[103]

In other words one should meaningfully speak of hermeneutics whose presence cannot be merely exorcised by invoking the notion of unmediated objectivity. Such a position as stated earlier does not necessarily negate the presence of an external reality independent of the human subject, but only states that our perception of this reality *qua* reality will be mediated by our linguistic and cultural perceptions as well. The filters we bring in become the frameworks with which we investigate the world.

With Symbolic AI, the framework with which researchers operate is that the mind is a machine. This framework is imposed to provide some sense of certainty and order to the phenomena being investigated. Herbert A. Simon himself acknowledges that such an imposition is necessary:

> Research on human thinking has progressed from relatively simple and well-structured phenomena (e.g. rote verbal learning, solving puzzles, simple concept attainment) to more complex and rather ill-structured tasks (e.g. use of natural language, learning, scientific discovery, visual art). 'Ill-structured' means that the task has illdefined or multidimensional goals, that its frame of

> reference or representation is not clear or obvious, that there are no clear-cut procedures for generating search paths or evaluating them – or some combination of these characteristics...When a problem is ill-structured in one or more of the senses, a first step in solving *it is to impose some kind of structure that allows it to be represented – that is, symbolized – at least approximately and attacked in this symbolized form* (emphasis mine).[104]

In Simon's case, the framework he is referring to is the notion that humans and machines are both physical symbol systems. Simon and his collaborator Newell hypothesized that strings of bit or symbols manipulated by a digital computer could stand for virtually any intelligent behavior. Newell explains this notion as follows:

> The digital-computer field defined computers as machines that manipulated numbers. The great thing was, adherents said, that everything could be encoded in numbers, even instructions. *In contrast, the scientists in AI saw computers as machines that manipulated symbols. The great thing was, they said, that everything could be encoded into symbols, even numbers* (emphasis mine).[105]

Coming up with a universal symbol system or a comprehensive programmable code for all human behavior would resolve the classical problem of mathesis (or formalization of everything) that Leibniz and others tried to articulate. The formulation of a symbol system as used in computer programming is closest to what Leibniz, Boole and others have along theorized about. Newell and Simon's symbol system hypothesis is based on the following assumptions:

> A physical symbol system consists of a set of entities, called symbols, which are physical patterns that occur as components of another type of entity called an expression (or symbol structure). Thus a symbol structure is composed of a number of instances (or tokens) of symbols related in some physical way (such as one token being next to another). At any instant of time the system will contain a collection of these symbol structures. Besides these structures, the system also contains a collection of processes that operate on expressions to produce other expressions...A physical symbol system is a machine that produces through time an evolving collection of symbol structures. Such a

> system exists in a world of objects wider than just these symbolic expressions themselves.[106]

Newell and Simon theorize that these symbols not only designate and interpret a larger body of symbols and object, but that they indeed constitute the basis for general intelligent action although there isn't necessarily any single elementary principle that accounts for intelligence in all its manifestations: "But the lack of a simple *deus ex machina* does not imply that there are no structural requirements for intelligence. One such requirement is the ability to store and manipulate symbols."[107] Newell and Simon suggest that symbol manipulation lies at the heart of intelligent action and that humans are also physical symbol systems. By the same token, machines are also physical symbol systems. Therefore, one could say that humans are equivalent to machines in the same manner in which we could say that A=B, based on the preceding assumptions A=C and B=C (where A stands for intelligent behavior in humans, B for intelligent behavior in machines, and C for general intelligent behavior by symbol-manipulation).

Newell and Simon state that the intellectual ancestry of their ideas can be traced to developments in mathematical logic, Turing's conception of universal Turing machines, the idea of programs that could store data and McCarthy's LISP (a high-level programming language that processes data in the form of lists that held data, with the lists serving as addresses that permitted access to other lists).[108] All the influences listed above involved the manipulation of symbols at some level or the other. Although the symbol system hypothesis was very instrumental in developing a lot of useful innovations like the Logic Theorist, the General Problem Solver and other expert systems, Simon and other symbolists overpredicted the scope of symbolic AI while underestimating the difficulty in formalizing common sense and real world knowledge. While these expert

> systems are really good at solving complex mathematical feats that is way past the average human mind, they are still far ways from performing really simple real world tasks that humans normally take for granted.[109]

Symbolic AI researchers chiefly rely on logical rules to execute their program. McCarthy's description of the role of logic in computer programs is illustrative:

> A machine on the lowest level uses no logical sentences. It merely executes the commands of its program...
> The next level of logic involves computer programs that put sentences in machine memory to represent their beliefs but use rules other than ordinary logical inference to reach conclusions...the sentences that appear in machine memory are from a program-dependent subset of the logical language being used...
> The third level uses first-order logic as well as logical deduction...
> It involves representing general facts about the world as logical sentences. Once put in a database, the facts can be used by any program...[110]

However, at the fourth level one is likely to encounter difficulties with regards to representing real world knowledge. Much of traditional logic has revolved around monotonic reasoning where all the premises are clearly stated, making the job of arriving at a conclusion much easier. However, most human conversations are nonmonotonic where all the premises are not fully stated. McCarthy states that more work in formalized "nonmonotonic reasoning" will come a long way in helping represent common sense or real world knowledge.[111]

On these grounds, it is unnecessary to write a hasty epitaph for symbolic AI. Neurology tells us that the human brain does not function like a computer program – however, this does not necessarily mean that symbolic AI should be abandoned.

Artificial Intelligence need not always rely on models of human intelligence. In fact, more work in symbolic AI should be encouraged even while welcoming alternative ways of approaching AI such as Connectionism and others.

The Mechanistic Metaphor as a Terministic Screen

Richard Boyd articulates eloquently that the 'mind is a computer' metaphor also gave rise to a number of generative concepts such as information-processing, encoding, decoding, feedback, memory stores among others that played an influential role in computational psychology:

> 1. the claim that thought is a kind of "information processing" and that the brain is a sort of "computer."
> 2. the suggestion that certain motoric or cognitive processes are "preprogrammed."
> 3. disputes over the issue of the existence of an internal "brain-language" in which "computations" are carried out.
> 4. the suggestion that certain information is "encoded" or "indexed" in "memory store" by "labeling," whereas other information is "stored" in "images."
> 5. disputes about the extent to which developmental "stages" are produced by the maturation of new "preprogrammed" "subroutines," as opposed to the acquisition of learned "heuristic routines," or the development of greater "memory storage capacities" or better "information retrieval procedures."
> 6. the view that learning is an adaptive response of a "self-organizing"machine.
> 7. the view that consciousness is a "feedback" phenomenon.[112]

Boyd articulates that some metaphors are "theory constitutive" in that they play much more than an exegetical role and help construct some of the salient theoretical concepts in a particular framework, engendered by the principal metaphor.

Out of the mind is a machine metaphor comes a whole conceptual vocabulary, the most obvious ones being:

> The brain is *hardware*
> The brain is a *rapid, complex calculating machine*

The brain is made up of *digital switches*
The mind is *software*
The mind is a *program* or *set of programs*
The mind *manipulates symbolic representations*
The mind is an *information machine*
Thinking is *computation*
Perception is *computation*
Memory is *looking up stored data*
The function of the mind and brain is *information-processing.*[113]

The computer is the source domain and the human mind is the target domain, where a computational vocabulary is mapped onto the human mind. Essentially, there is a sense that if an isomorphic mapping can occur, the mind should be seen as a mathematical, problem space and intelligence a product of mental operations carried out within that space. By viewing the mind as a problem-space, two concurrent ideas emerge namely one of problem-solving in a computational sense and the other, of immense space where a whole range of mental activities are translated in a computational terms. However the machine metaphor in symbolic AI is a little more elevating than behaviorist psychology's metaphor of mind, which goes as well – animals are reflex machines, humans are animals and therefore humans are also reflex machines. Paradigmatic experiments in behaviorist psychology, be it Pavlov's dogs, Tolman's rats or Skinner's pigeons make an association between a natural reflex (salivation at the sight or smell of food) and an arbitrary stimulus (such as the sound of a bell). By the same token, if humans are reflex machines even the mental process by which we associate words such as "dinnertime" with food is considered a purely a reflexive, stimuli-response process.

According to Paul N. Edwards, the reflex machine metaphor has certain parallels with the computer metaphor but leads in completely different directions: "For example, symbolic activity (such as language, problem-solving, and perception), physical behavior, and emotional responses are all on a par under the reflex machine metaphor. The metaphor directs attention toward the external variables controlling a response rather than toward the structure of (complex) established behavior patterns."[114] However, the points of divergences are as follows:

> If the mind is a computer, it may be reprogrammed, while if it is a reflex machine, its responses may be modified through new conditioning. While reprogramming and behavior modification are different processes, they have in common the precept of a flexibility of the mental apparatus and the possibility of change and learning...The reflex machine metaphor concentrates...on environmental variables as triggers for behavior, suggesting a focus on the social system of rewards as the ultimate "technology" of behavior...The computer metaphor instead draws attention to the internal structure of the mind and its representational schemes. It suggests the possibility of "reprogramming" the mind by setting up new thoughts patterns or restructuring its "hardware" with drugs, surgery, or implanted microchips.[115]

An obvious protestation against the mechanistic reduction of mental processes to programmable procedures is to say that there are qualitative aspects of human experience as well. However, the programmer's comeback is best expressed in Herbert Simon's remarks, who makes a distinction between "well-structured" and "ill-structured" phenomena – well-structured phenomena pertain to clearly defined domains such as problem-solving, rote verbal learning among others, while ill-structured phenomena pertains to "use of natural language, learning, scientific discovery, visual art" among others.[116] Simon suggests "when a problem is ill-structured," a first step in attacking the problem is "to impose some kind of structure that allows it to be represented – that is symbolized – at least approximately."[117] The top-down approach in the sciences is considered ubiquitous if one were to make scientific progress:

> We knew a great deal about the gross physical and chemical behavior of matter before we had a knowledge of molecules, a great deal about molecular chemistry before we had an atomic theory, and a great deal about atoms before we had any theory of elementary particles...The skyhook-skyscraper construction of science from the roof down to the yet unconstructed foundations was possible because the behavior of the system at each level depended on only a very approximate, simplified, abstracted characterization of the system beneath.[118]

The top-down approach is quite evident in symbolic AI, starting with a strong metaphor that creates more subcategories within the computational paradigm. THE MIND IS A COMPUTER WITH DISCRETE STATES WHAT GOES ON IN THE MIND ARE MENTAL PROCESSES WHICH ARE SEEN AS SYMBOL-MANIPULATION – Therefore, the mind is a symbol processor. Symbol-manipulation or thought involves step-by-step ALGORITHMIC PROCEDURES. These procedures involve routines and subroutines.

All the above-mentioned examples of how the computer metaphor plays out leads us back to the notion that the terministic screen engenders - namely, the mind is a mathematical, problem space. The idea that the mind is a problem-space chiefly construes the mind as a mathematical space where thought is not only computation, but also thought is also a formal process that revolves around some notion of formalization of knowledge. In order to think of the mind as a problem space, it is important to consider all mental processes as tacit physical behaviors. These physical behaviors have certain cognitive structures. These cognitive structures in turn can be represented as frames:

> A frame is a data structure for representing a stereotyped situation, like being in a certain kind of living room, or going to a child's birthday party. Attached to each frame are several kinds of information. Some of this information is about how to use the frame. Some is about what one can expect top happen next. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed. We can think of a frame as a network of nodes and relations. The top levels of a frame are fixed, and represent things that are always true about the supposed situation. The lower levels have many terminals – slots that must be filled by specific instances or data. Each terminal can specify conditions its assignments must meet.[119]

The cognitive structures of the mind are descriptive, data structures representable as frames. If this is the case, even qualitative content is represented by means of frames. Minsky goes on to describe four types of dominant frames that occur in normal conversational settings:

> Surface syntactic frames – mainly verb and noun structures. Prepositional and word-order indicator conventions
> Surface semantic frames – action-centered meanings of words , qualifiers and relations concerning participants, instruments, trajectories and strategies, goals, consequences, and side-effects.
> Thematic frames – Scenarios concerned with topics, activities, portraits, settings.
> Narrative frames – Skeleton forms for typical stories, explanations, and arguments. Conventions about foci, protagonists, plot arrangements, plot forms, development, and so on, designed to help a listener construct a new, instantiated frame in his own mind.[120]

These frames are in turn indicative of a wide-range of mental activities and situations, all of which are contingent on the formulation of systematic rules of logic. Different situations are represented in terms of identifiable objects possessing properties that are well defined.

The computational metaphor which engenders the view of the mind as a mathematical, problem space makes us focus our attention in particular ways by mapping on the linguistic resources of the source domain onto the target domain. Mapping is said to occur even for complex types of activities, once the activity is scripted or represented as a frame. Given the fact, that mapping do indeed shed insight on the psychology of mind, from a computational perspective – a related question that one should ask is – 'what is outside the frame?' Frames in their search for a scripted concreteness of particular situations between mental actors tend to ignore the areas that are outside the frame. These include areas pertaining to intuition and non-formalizable mental behavior. What is rhetorically significant is the hope generated by defining the mind as a mathematical, problem space. The immensity that space suggests simply entails that a whole range of problems or mental activities are considered computational and thinking is largely considered problem solving.

The other omissions include a neglect of motivation – for instance, a frame in itself does not comment on the motivation behind the frame, the intent behind preferring one script over another or why the script is there in the first place. A close analogy that comes to mind is the idea of a proof in a mathematical theorem – the theorem by itself does not understand the concept of a proof, the mathematician intuitively has in his or her head some notion of why a proof is important. The step-by-step process of working out the proof is an internal motivation, which is not disclosed in the actual theorem. However, on the other hand, thinking of oneself in computational terms does have distinct advantages:

> The metaphor of the mind as a set of programs, or symbolic instructions that process inputs and control outputs, provides a rich set of analogies that allow us to portray the complex, hidden, abstract processes of thinking and the production of behavior in terms of the relatively simpler and more concrete ones involved in computer programming. Like much human behavior, most computer programs are not built in or "hard-wired." This implies that behavior and thought patters can be changed, erased or replaced. Imperfect computer programs have "bugs" – flawed instructions that cause erratic, unwanted results. Human behavior and thought, too, can "go haywire." The computer metaphor implies that with diligence bugs can be located and corrected.[121]

Yet, what is largely left out from the computational metaphor (based on our existing technology) is the role that emotions play in shaping and influencing thought. For instance, even the very desire to be rational is intricately intertwined with an emotional dispositions such as 'balance,' 'stability' and so forth. Even if there is scope for accommodating emotions, they are to be primarily represented as a symbolic and mathematical process. Thus in the very aspect of representation, some of the phenomenological and qualitative aspects of emotions are lost. This is not to suggest that one should not study emotions from a computational perspective, but only to describe that our metaphoric models of mind are only partial and not complete because something is always left out. Edwards explains:

> The computer metaphor also implies, then, that emotion is either irrelevant to the understanding of human thought, or that emotion might somehow be represented as a symbolic process. The computer is a logic machine. Thus the computer metaphor privileges one mode of human thought at the expense of other, paralogical or tropological modalities. It points toward a reductive explanation of the paralogical, the tropological, and the intuitive in terms of a more rigorous, mathematical or quasi-mathematical logic.[122]

The pervasive power and persuasive appeal of the computer metaphor is manifest in linguistic practices (at least in the English language) where people refer to their brains as being turned 'on' or 'off,' stubborn habits or ingrained tendencies are 'hard-wired,' 'programmed' and so forth, thinking and reflection are considered 'information-processing,' memory is referred to as 'stored' 'data,' incoming information to the brain or to the eye are seen as 'inputs,' adaptive behavior as a result of the 'processed information' are considered 'outputs,' too much information is considered 'information overload,' and the list goes on.

As stated earlier, the mind is a machine metaphor is a pervasive and powerful one in computational psychology, with its latent and manifest power residing in doing away with the distinction between literal and conveyed meanings. Pylyshyn's position elucidates the reason for why an isomorphism between humans and machines insofar as mental activity is concerned is necessary:

> It seems to me that *computation*, and all that it entails regarding rule-governed transformations on intentionally interpreted symbolic expressions, applies just as literally to mental activity as it does to the activity of digital computers. Such a term is in no sense a literal description of the operation of electronic computers that has been metaphorically transported to the primary subject of mind. The relation between computation and artifacts is just as abstract as the relation between computation and mental activity...Both require that we give terms an interpretation in these domains, precisely the way Euclidean axioms are given a realistic interpretation in classic space, and the axioms of non-Euclidean geometry are given a realistic interpretation in the space of special relativity. I see no significant differences here which would lead me to characterize computation, but not geometry, as metaphor.[123]

Quite clearly, Pylyshyn seems to be arguing that the computational 'metaphor' (although he evidently prefers to think of it as a literal description) should be seen as no less than a representation of mental activity. Computational descriptions are seen as geographic metaphors or maps of mind. Two questions that are crucial to computational psychologists are "what is the relation of intelligence to the internal world of the individual?" and "what are the information-processing routines (programs) underlying intelligent thought?"[124]

A natural question may arise at this juncture, pertaining to the literality or figurative aspects of computational paradigm. What aspects are literal, and what aspects are figurative are metaphorical exactly are hard to tell. But needless to say, the computational metaphor has created conceptual schemas and components with which to view human or nonhuman intelligences. Other interrelated metaphoric approaches worth noting are as follows.

Robert J. Sternberg identified the following features emerging from the computational metaphor:

> a) Processes that include algorithmic and heuristic strategies.
> b) knowledge (expert systems, knowledge base engineering pertaining to the representation of knowledge, knowledge organization pertaining to how complex storage systems find and access prior knowledge and structure its knowledge in such a way that it generates behavior towards problem solving, beliefs which are very high-level knowledge structures that can influence the ability of the system to comprehend, inferences pertaining to the ability of high-level knowledge structures that can aid in the processing of familiar information, architecture, levels of understanding and Learning.[125]

The computational paradigm suggests that mental activity takes place in discrete mental states. The mental states evoked by the computational metaphor can be further seen as an "event structure metaphor," where:

> States are locations (bounded regions in space)
> Changes are movements (into or bounded regions)
> Causes are forces.
> Actions are self-propelled movements.
> Purposes are destinations.
> Means are paths (to destinations)
> Difficulties are impediments to motion.

> Expected progress is a travel schedule; a schedule is a virtual traveler, who reaches prearranged destinations at prearranged times.
> External events are large, moving objects.
> Long term, purposeful activities are journeys.[126]

Overall, as Lakoff explains "the metaphor is not just a matter of language, but of thought and reason. The language is secondary. The mapping is primary, in that it sanctions the use of source domain languages and inference patterns for target domain concepts."[127] Boyd ruminates that the computational metaphor has been influential in generating theoretical terms that engage in reference fixing:

> The case of computer metaphors in cognitive psychology, I believe, illustrates this sort of ostensive introduction of theoretical terminology. Mental and psychological states and processes are, almost certainly, among the sorts of kinds whose essential properties are relational – they are functional states or processes...Theoretical terms in psychology, then, are among those for which reference fixing must typically involve disambiguation between several quite different but co-occuring kinds of the same sort. Computational states and processes...are also functional states and processes: typically their essential properties are their causal relations to other computational states or processes or to the inputs and outputs of the machines which realize or manifest them. What I am suggesting is that – when computer metaphors in cognitive psychology are successful – the metaphorically employed computer terms come to have new referents in the context of psychological theory construction. They refer to functionally defined psychological states or processes which bear to each other functional relations analogous to those which the literal referents of these terms bear to one another. If the metaphors are apt, and if they are drawn in sufficient detail, the difference in functional (relational) properties of the literal referents of the computer terms will serve – by analogy – to disambiguate the referents of these terms in their theory-constitutive metaphorical applications.[128]

Boyd's contention is well-taken, although it is debatable whether one could completely disambiguate referents from a chain of descriptions, especially in instances where subject-predicate relationships are inverted and blurred. The primary formulation 'the mind is a computer' could very well be reversed with the 'computer is a mind' in computational psychology. Given the conflation of distinctions, maintaining composure of disambiguation in the face of constructivism is perhaps questionable. Richard E. Mayer's postulation that metaphors are instructive, given that they enhance understanding about a subject by providing new mental concepts is heuristic:

> Constructivism is based on the idea that human understanding is the result of mental construction by the learner – it is based on the learning-as-construction metaphor. According to the constructivist approach, the language of science serves to help people construct an understanding of science. For example, to help understand the relation between heat and volume, students may need to construct the concept of particles in a box.[129]

In effect, the mental construction in computational psychology deals with treating the mind as a computing machine: "cognitive science...tries to elucidate the workings of the mind by treating them as computations."[130]

As discussed in the preceding chapters, there are distinct advantages to the computational paradigm. Such a paradigm clearly articulates step-by-step the algorithm like procedures of mental operations. Empirical verifiability, consistency, replicability and a sense of concreteness are accorded to the psychological enterprise that was perhaps lacking in pre-computational models. On the other hand, there are significant rhetorical elisions in such a framework, that noncomputable features if any are either explained away or considered irrelevant. The problems with the classic determinist models of reality resurface in the determinism of computational paradigms. The active, constructivist roles that human agents play in building up the edifices of cumulative scientific knowledge is significantly downplayed; furthermore from a metacommunicative perspective the linguistic framing problem is ignored and competing explanations are dismissed beforehand. The technical precision and the subsequent impact of effective results ostensibly lead computationalists to reject alternative explanations. Thomas Kuhn's formulation that one could arrive at empirical results using different paradigms is perhaps not heeded. Henri Poincaré illustrates how particular frameworks generate specific ways of seeing, which also become ways of not seeing in other ways:

> Beings whose minds were made as ours, and with senses like ours, but without any preliminary education, might receive from a suitably-chosen external world impressions which would lead them to construct a geometry other than that of Euclid, and to localize the phenomena of this external world in a non-Euclidean space, or even in space of four dimensions. As for us, whose education has been made by our actual world, if we were suddenly transported into this new world, we should have no difficulty in referring phenomena to our Euclidean space. Perhaps somebody may appear on the scene some day who will devote to his life to it, and be able to represent to himself the fourth dimension...If geometrical space were a framework imposed on each of our representations considered individually, it would be impossible to represent to ourselves an image without this framework, and we should be quite unable to change our geometry. But this is not the case; geometry is only the summary of the laws by which these images succeed each other. There is nothing, therefore, to prevent us from imagining a series of representations, similar in every way to our ordinary representations, but succeeding one another according to laws which differ from those to which we are accustomed. We may thus conceive that beings whose education has taken place in a medium in which those laws would be so different, might have a very different geometry from ours.[131]

What is true for geometry could be said for computation as well. What if, one had a different conception of computation quite unlike the models present in symbolic AI and computational psychology? And if that conception of computation were empirically verifiable, the nature of mind, according to this new school thought would be essentially different. Therefore, the issue is not so much whether the computational metaphor is true or false insofar as representation is concerned, instead the issue is whether the paradigm is heuristic, from an explanatory perspective or not. With that regard, the existing computational metaphor has been more than useful for cognitive scientists who find intellectual affinity with this mode of thinking.

Notes

[1] Boethius, *Consolation of Philosophy*, trans. V.E.Watts (London: The Folio Society, 1998), 167.
[2] Hans Moravec, *Mind Children* (Cambridge, Massachusetts: Harvard University Press, 1988), 1-2.
[3] Marvin Minsky, *The Society of Mind* (New York: Simon and Schuster, 1985), 17.
[4] James Crosswhite, *The Rhetoric of Reason: Writing and the attractions of argument* (Madison, Wisconsin: The University of Wisconsin Press, 1996), 162.
[5] Crosswhite, *The Rhetoric of Reason*, 162.
[6] John Haugeland, *Mind Design II: Philosophy, Psychology, Artificial Intelligence* (Cambridge, Massachusetts: Bradford Book, MIT Press), 19-20.
[7] Haugeland, *Mind Design II*, 16.
[8] A. M. Turing, "Lecture to the London Mathematical Society on 20 February, 1947" in *A.M.Turing's ACE Report of 1946 and Other Papers* (London: Charles Babbage Institute Reprint Series for the History of Computing), 106-124. (112).
[9] Turing, "Lecture to the London Mathematical Society," 122.
[10] Turing, "Lecture to the London Mathematical Society," 122.
[11] Herbert Simon, "Machine as Mind," in *Machines and Thought: The Legacy of Alan Turing, Vol. I*, ed. P. J. R. Millican and A. Clark (Oxford: Clarendon Press, 1996), 81-102 (82).
[12] Herbert A. Simon, "Thinking by Computers," in *Mind and Cosmos: Essays in Contemporary Science and Philosophy*, ed. Robert. G. Colodny (Pittsburgh: The University of Pittsburgh Press, 1966), 18.
[13] Allen Newell and Herbert A. Simon, "Computer Science as Empirical Inquiry: Symbols and Search," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: Bradford Books, MIT Press), 81-110 (cited in page 87)
[14] Newell and Simon, "Computer Science as Empirical Inquiry," 85.
[15] Newell and Simon, "Computer Science as Empirical Inquiry," 86.
[16] Newell and Simon, "Computer Science as Empirical Inquiry," 84.
[17] Newell and Simon, "Computer Science as Empirical Inquiry," 87.
[18] Paul Smolensky, "Connectionist Modeling: Neural Computation/Mental Connections," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: Bradford Books, MIT Press), 233-250 (cited in page 235).
[19] Newell and Simon, "Computer Science as Empirical Inquiry."
[20] Philip E. Agre, *Computation and Human Experience* (Cambridge: Cambridge University Press, 1997), 51.
[21] Agre, *Computation and Human Experience*, 51.
[22] Simon, "Machine as Mind," 83.
[23] Simon, "Machine as Mind," 83.
[24] Frank Rosenblatt, "Strategic Approaches to the Study of Brain Models," in *Principles of Self-Organization*, ed. H. von Foerster (Elmsford, New York: Pergamon Press, 1962).
[25] Paul N. Edwards, *The Closed World: Computers and the Politics of Discourse in Cold War America* (Cambridge, MA: MIT Press, 1996), 34.
[26] George Lakoff and Mark Johnson, *Metaphors We Live By* (Chicago: The University of Chicago Press, 1980), 11-12.
[27] Noam Chomsky, *Aspects of the Theory of Syntax* (Cambridge, Massachusetts: MIT Press, 1965)
[28] Garrett Birkhoff, "Mathematics and Psychology," in *SIAM Review* 11(October, 1969): 429-469 (cited in 440)

[29] Hubert L. Dreyfus and Stuart E. Dreyfus, *Mind over Machine: The Power of Human Intuition and Expertise in the era of the computer* (New York: Free Press, 1986), 54.
[30] Agre, *Computation and Human Experience,* 87.
[31] Lakoff and Johnson, *Metaphors We Live By,* 36.
[32] Agre, *Computation and Human Experience,* 46-47.
[33] Simon, "Machine as Mind," 88.
[34] cited in Paul Tunbridge, *Lord Kelvin, his influence on electrical measurements and units* (London: Peter Peregrinus Ltd on behalf of the Institution of Electrical Engineers, 1992), 17-18.
[35] cited in Crosbie Smith and M. Norton Wise, *Energy and Empire: A Biographical Study of Lord Kelvin* (Cambridge: Cambridge University Press, 1989), 149.
[36] Albert Einstein, "Maxwell's Influence on the Development of the Conception of Physical Reality," in *James Clark Maxwell: A Commemoration Volume 1831-1931* (Essays by Sir J. J. Thomson, Max Planck, Albert Einstein, Sir Joseph Larmor, Sir James Jeans, William Garnett, Sir Ambrose Fleming, Sir Oliver Lodge, Sir R. T. Glazebrook, Sir Horace Lamb) (Cambridge: Cambridge University Press, 1931), 67-68.
[37] L. M. Laita, "Influences on Boole's logic: The controversy between William Hamilton and Augustus de Morgan," *Ann. Sci* 36 (1979): 45-65.
[38] George Boole, *The mathematical analysis of logic: being an essay towards a calculus of deductive reasoning* (Cambridge: Cambridge University Press, 1847), pp. 9-13.
[39] George Boole, "Logic," in *George Boole: Selected Manuscripts on Logic and its Philosophy,* Ed. Ivor Grattan-Guinness and Gerard Bornett (Boston: Birkhäuser Verlag, 1997), 126-153 (126)
[40] George Boole, *An Investigation of the Laws of Thought: on which are founded The Mathematical Theories of Logic and Probability* (New York: Dover, 1854), 45-46.
[41] Boole, *An Investigation of the Laws of Thought,* 1.
[42] Boole, *An Investigation of the Laws of Thought,* 4.
[43] Boole, *An Investigation of the Laws of Thought,* 5.
[44] Boole, *An Investigation into the Laws of Thought,* 6.
[45] Boole, *An Investigation into the Laws of Thought,* 6.
[46] George Boole, "On the Foundations of the Mathematical Theory of Logic," in *George Boole: Selected Manuscripts on Logic and its Philosophy,* Ed. Ivor Grattan-Guinness and Gerard Bornett (Boston: Birkhäuser Verlag, 1997), 63-104 (67).
[47] Boole, "On the Foundations of the Mathematical Theory of Logic," 86.
[48] Boole, "On the Foundations of the Mathematical Theory of Logic," 86-87.
[49] Boole, "On the Foundation of the Mathematical Theory of Logic," 87.
[49] Boole, "On the Foundation of the Mathematical Theory of Logic," 87
[50] Boole, "On the Foundation of the Mathematical Theory of Logic," 88.
[51] Boole, "On the Foundation of the Mathematical Theory of Logic," 88.
[52] Boole, "On the Foundation of the Mathematical Theory of Logic," 88-89.
[53] W. V. Quine, *Selected Logic Papers* (New York: Random House, 1966), 1.
[54] C. I. Lewis, *A Survey of Symbolic Logic* (Berkeley: University of California Press, 1918), 51.
[55] Boole, "On the Foundation of the Mathematical Theory of Logic," 89.
[56] Boole, "On the Foundation of the Mathematical Theory of Logic," 89-90.
[57] Boole, "On the Foundation of the Mathematical Theory of Logic," 91.
[58] Richard Whately, *Elements of Logic* (Boston: James Munroe and Company, 1859), 19.
[59] Agre, *Computation and Human Experience,* 89.
[60] Desmond MacHale, *George Boole: His Life and Work* (Dublin: Boole Press Limited, 1985), 71-72.
[61] A.M. Turing, "On Computable Numbers, With an application to the Entscheidungsproblem," *Proceedings London Mathematical Society* Series 2, 42 (1936): 230-265 (250).
[62] Chris Fields, "Measurement and Computational Description," in *Machines and Thought: The Legacy of Alan Turing, Vol. I,* ed. P. J. R. Millican and A. Clark (Oxford: Clarendon Press, 1996),165-177 (176).
[63] Turing, "On Computable Numbers," 231.
[64] Turing, "On Computable Numbers," 232.

[65] Turing, "On Computable Numbers," 253-254.
[66] Aaron Sloman, "Beyond Turing Equivalence," in *Machines and Thought: The Legacy of Alan Turing, Vol. I*, ed. P. J. R. Millican and A. Clark (Oxford: Clarendon Press, 1996), 179-219.
[67] Sloman, "Beyond Turing Equivalence," 181.
[68] Sloman, "Beyond Turing Equivalence," 184.
[69] cited in Anthony Galton, *The Church-Turing Thesis: Its Nature and Status* in *Machines and Thought: The Legacy of Alan Turing, Vol. I*, ed. P. J. R. Millican and A. Clark (Oxford: Clarendon Press, 1996), 137-164 (p. 146).
[70] Roger Penrose, *The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics* (New York: Penguin Books, 1989), 432-433.
[71] Penrose, *The Emperor's New Mind*, 448.
[72] John R. Searle, "Minds, Brains, and Programs," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: Bradford Books, MIT Press, 1997), 183-204.
[73] John McCarthy, "Programs with Common sense," in *Formalizing Common Sense: Papers by John McCarthy*, ed. Vladimir Lifschitz (Norwood, New Jersey: Ablex Publishing Corporation, 1990), 9-16 (9).
[74] John McCarthy and P. J. Hayes, "Some Philosophical Problems From the Standpoint of Artificial Intelligence," in *Formalizing Common Sense: Papers by John McCarthy*, ed. Vladimir Lifschitz (Norwood, New Jersey: Ablex Publishing Corporation, 1990), 21-63 (25).
[75] McCarthy and Hayes, "Some Philosophical Problems From the Standpoint of Artificial Intelligence," 26.
[76] John McCarthy, "Ascribing Mental Qualities to Machines," in *Formalizing Common Sense: Papers by John McCarthy*, ed. Vladimir Lifschitz (Norwood, New Jersey: Ablex Publishing Corporation, 1990), 93-118 (93).
[77] John McCarthy, "Mathematical Logic in Artificial Intelligence," in *The Artificial Intelligence Debate: False Starts, Real Foundations*, ed. Stephen R. Graubard (Cambridge, Massachusetts: The MIT Press, 1988), 297-311.
[78] McCarthy, "Ascribing Mental Qualities to Machines," 95.
[79] Ukrainian Chess Online, "Kasparov-Deep Blue, 1997," http://chess-sector.odessa.ua/kas_deep.htm
[80] Vishwanathan Anand, "More questions than answers," http://www.research.ibm.com/deepblue/home/may11/story_3.html
[83] Keith J. Holyoak and Paul Thagard, *Mental Leaps – Analogy in Creative Thought* (Cambridge, Massachusetts: The MIT Press, 1995), p. 211.
[84] Kenneth J. Knoepsel, "The Narrative Matter of Mathematics: John Dee's Preface to the *Elements* of Euclid of Megara," *Philosophical Quarterly* 66(1987): 27-46 (42).
[85] René Descartes, "Meditations II, VI," in *Problems in Mind: Readings in Contemporary Philosophy of Mind*, ed. Jack S. Crumley II (Mountain View, California: Mayfield Press, 2000), 21-33 (28).
[86] Thomas Hobbes, "Leviathan (Selections)," in *Classics of Modern Political Theory: Machiavelli to Mill*, ed. Steven M. Cahn (New York: Oxford University Press, 1997), 78-196 (80).
[87] A. P. Martinich, "Thomas Hobbes," in *Modern Philosophers: From Descartes to Nietzsche*, ed. Steven M. Emmanuel (Oxford: Blackwell, 2001), 28-42 (31).
[88] Hobbes, "Leviathan," 86-87.
[89] Gottfried Wilhelm Leibniz, *Theodicy*, ed. Diogenes Allen (New York: Bobbs-Merrill, 1966), 51.
[90] Cited in Margaret Dauler Wilson, *Ideas and Mechanism: Essays on Modern Philosophy* (Princeton, New Jersey: Princeton University Press, 1999), 391.
[91] Cited in Wilson, *Ideas and Mechanism*, 390.
[92] Margaret Dauler Wilson, "Leibniz and Materialism," in *Ideas and Mechanism: Essays on Modern Philosophy*, ed. Margaret Dauler Wilson (Princeton, New Jersey: Princeton University Press, 1999), 388-406 (390).
[93] G. W. Leibniz, "New System of the nature of substances and their communication, and of the union which exists between the soul and the body," in *G. W. Leibniz: Philosophical Texts*, ed. R. S. Woolhouse and Richard Franks (New York: Oxford University Press, 1998), 143-152 (151).

[94] G. W. Leibniz, "Monadology," in *G. W. Leibniz: Philosophical Texts*, ed. R. S. Woolhouse and Richard Franks (New York: Oxford University Press, 1998), 266-281 (266-277).
[95] Cited in François Dagognet, "Materialism: A Philosophy of Multiple Revivals," *Graduate Faculty Philosophy Journal* 22 (2000): 3-14 (page 5).
[96] Denis Diderot, *Oeuvres Complès*, ed. Roger Lewinter (Paris: Club Français du Livre, 1969-1973), 951.
[97] Julien Offray de La Mettrie, *Machine Man and Other Writings*, trans. Ann Thomson (Cambridge: Cambridge University Press, 1966), 66.
[98] Ann Thomson, "La Mettrie, Machines, and the Denial of Liberty," *Graduate Faculty Philosophy Journal* 22 (2000): 71-86 (73).
[99] La Mettrie, *Machine Man and Other Writings*, 26.
[100] La Mettrie, *Machine Man and Other Writings*, 39.
[101] Joseph Weizenbaum, *Computer Power and Human Reason: From Judgment to Calculation* (San Francisco: W. H. Freeman and Company, 1976), 21.
[102] Alan Gross, "On the Shoulders of Giants: Seventeenth-century optics as an argument field," *Quarterly Journal of Speech* 74 (1988): 1-17 (13).
[103] Charles Alan Taylor, *Defining Science: A Rhetoric of Demarcation* (Madison: The University of Wisconsin Press, 1996), 113.
[104] Simon, "Machine as Mind," 93.
[105] cited in Hubert L. Dreyfus and Stuart E. Dreyfus, "Making a Mind Versus Modeling the Brain: Artificial Intelligence Back at a Branchpoint," in *The Artificial Intelligence Debate: False Starts, Real Foundations*, ed. Stephen R. Graubard (Cambridge, Massachusetts: The MIT Press, 1988), 15-43 (3).
[106] Newell and Simon, "Computer Science as Empirical Inquiry," 85-86.
[107] Newell and Simon. "Computer Science as Empirical Inquiry," 83.
[108] Newell and Simon, "Computer Science as Empirical Inquiry," 88-91.
[109] David Willshaw, "Non-Symbolic Approaches to artificial intelligence and the mind," *Philosophical Transactions: Physical Sciences and Engineering* 349 (1994): 87-101.
[110] McCarthy, "Mathematical Logic In Artificial Intelligence," 299-303.
[111] McCarthy, "Mathematical Logic in Artificial Intelligence."
[112] Boyd, "Metaphor and theory change," 486.

[113] Paul N. Edwards, *Computers and the Politics of Discourse in Cold War America* (Cambridge, Massachusetts: MIT Press, 1996), 161.
[114] Edwards, *Computers and the Politics of Discourse in Cold War America*, 163.
[115] Edwards, *Computers and the Politics of Discourse in Cold War America*, 164.
[116] Herbert Simon, "Machine as Mind," in *Machines and Thought: The Legacy of Alan Turing, Vo. I*, eds. P. J. R. Millican and A. Clark (Oxford: Clarendon Press, 1996), 81-102 (91).
[117] Simon, "Machine as Mind," 91.
[118] Herbert A. Simon, *The Sciences of the Artificial* (Cambridge, Massachusetts: MIT Press, 1969), 17.
[119] Marvin Minsky, "A Framework for Representing Knowledge," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: Bradford Books, MIT Press), 111-142 (111=112).
[120] Minsky, "A Framework for Representing Knowledge," 123.
[121] Edwards, *Computers and the Politics of Discourse in Cold War America*, 162.
[122] Edwards, *Computers and the Politics of Discourse in Cold War America*, 162.
[123] Pylyshyn, "Metaphorical Imprecision and the "top-down" research strategy," 557.
[124] Robert J. Sternberg, *Metaphors of Mind: Conceptions of the Nature of Intelligence* (Cambridge: Cambridge University Press, 1990), 5.
[125] Sternberg, *Metaphors of Mind*, 130-140.
[126] George Lakoff, "The contemporary theory of metaphor," in *Metaphor and Thought, Second Edition*, ed. Andrew Ortony (Cambridge: Cambridge University Press, 1993), 202-251 (220).
[127] Lakoff, "The contemporary theory of metaphor," 208.
[128] Boyd, "Metaphor and theory change," 494.

[129] Richard E. Mayer, "The instructive metaphor: Metaphoric aids to students' understanding of science," in *Metaphor and Thought, Second Edition*, ed. Andrew Ortony (Cambridge: Cambridge University Press, 1993), 561-578 (563).
[130] P. N. Johnson-Laird, *The Computer and the mind: An Introduction to cognitive science* (Cambridge, Massachusetts: Harvard University Press, 1988), 9.
[131] Henri Poincaré, "Reading: Space and Geometry," in *Space from Zeno to Einstein: Classic Readings with a Contemporary Commentary*, ed. Nick Huggett (Cambridge, Massachusetts: MIT Press, 1999), 235-240 (233-235).

CHAPTER VI

FROM SYMBOLS TO NEURONS

The discussion in the preceding chapter attempted to rhetorically delineate the features of a thread, that wove a fabric etched with the letters 'the mind is a machine.' The rapid development of the personal computer enabled the extrapolation of this technology as a master trope or paradigm to guide the study the mind. Humanity's recurring fascination with technology enables a dialectical relationship between the user and the artifact, whereby the user vastly enhances his or her capabilities by engaging the artifact, and by the same token, the artifact provides a powerful and alternative means by which the user could reconceptualize the world of which he or she happens to be a part. As stated in preceding chapters, the invention of clocks, not inadvertently, coincided with clockwork accounts of the universe and by the same token, the invention of the computer coincides with computational accounts of physical reality.

The finite, problem space of a computational paradigm is rhetorically compelling in that it accords an undeniable sense of calculability and predictability combined with attempts to demystify anything that is considered 'unknowable.' Symbolic AI in effect best represents a programmer's model of intelligence in which thought is primarily symbol-manipulation, following a logico-mathematical route excelling at problem solving in task domains that demands mechanical reasoning. However, simulating tasks that are by far simpler to the human mind, especially tasks that take place in real-time environments have become imposing obstacles for symbolic modeling.

Connectionism proposes a way around this impasse by opting for a biologistic route in positing models that seek to be adequately representative of neuronal circuitry in the human brain. A perfunctory glance at connectionist literature shows that the 'mind is a machine' hypothesis derived from serial computing is replaced by an even more robustly organic thesis, namely that 'computers are like brains' hypothesis derived from the parallel processing brain. Thus instead of embracing a top-down paradigm, connectionists are predisposed to bottom-up approaches – namely, building better machines by studying an organic machine – the human brain. Set against this backdrop, the purpose of this chapter is two-fold: a) trace the rhetorical situation facilitating the emergence of connectionism, and b) undertake a rhetorical analysis of the assumptions of mind presupposed by connectionist models by highlighting both the rhetorical insights and blind spots.

Rhetorical Situation

The emergence of connectionist AI can be traced to at least three overlapping and influential intellectual trends: 1) nineteenth century experimental medicine and early behavioral psychology, 2) the appropriation of neural models in the emergence of parallel-distributed processing and the philosophical departure from symbolic AI and 3) the flourishing of the neurosciences with its increasing emphasis on eliminative materialist accounts of consciousness. All the above-mentioned trends display common currency insofar as signifying a shift from *a priori* logical and mathematical models to more realistic biological models of studying living systems.

Influences

Experimental physiology and early behavioral psychology play an important role in providing contextual coherence and rhetorical validation to the development of connectionism. A historical approach is helpful in tracing the developmental contours that enables the rhetorical acceptance of connectionist models of mind. One of the pioneering names who figures prominently in almost any history of systematic, empirical medicine is Claude Bernard (a prominent physiologist and a philosopher of science) who saw the human mind and body as a connection system.

As early as 1865, Bernard famously declared: "the study of life includes two things: (1) study of the properties of organized units; (2) study of the organic environment, i.e. study of the conditions which this environment must fulfill to permit the appearance of vital activities."[1] The philosophy of vitalism, as is well known, cast mental phenomena such as consciousness as irreducible complexities. Bernard extended the scope of physiological investigation to all aspects of biological activities including the mental, by seeking causal explanations between physico-chemical interconnections of tissues and their subsequent manifestations. Appraising his contributions towards enhancing a physiological understanding of living systems as follows, Bernard explains:

> I think I was the first to urge the belief that animals have really two environments: a *milieu extérieur* in which the organism is situated, and a *milieu intérieur* in which the tissue elements live. The living organism does not live in the *milieu extérieur* but in the liquid *milieu intérieur* formed by the

> circulating organic liquid which surrounds and bathes all tissue elements...the liquid part of the blood which, in the higher animals, is diffused through the tissues and forms the ensemble of the intercellular liquids and is the basis of all local nutrition and the common factor of all elementary exchanges. *A complex organism should be looked upon as an assemblage of simple organisms which are the anatomical elements that live in the liquid milieu intérieur.*[2]

In the *milieu intérieur* the presence of dynamic interconnections between these anatomical units giving rise to the appearance of individual phenomena, including vital phenomena:

> Differing living units thus play the part of stimuli, one in relation to another; and the functional manifestations of an organism are merely their harmonious reciprocal relations. The histological (the study of tissues) units react either separately or one against another by means of vital properties which are themselves in necessary connection with surrounding physico-chemical conditions...therefore...*we must not set up an antagonism between vital phenomena and physico-chemical phenomena, but on the contrary, we must note the complete and necessary parallelism between the two classes of phenomena...The role of men of science is to try to define and determine the material conditions producing the appearance of each phenomenon.*[3]

Vital phenomena are generated by physico-chemical material conditions although they may seem less apparent since "what we see from the outside is merely the result of physico-chemical stimuli from the inner environment; that is where physiologists must build up the real determinism of vital functions."[4] Furthermore, humans are seen as living machines that are organically self-regulated, in such a fashion that Bernard finds it "impossible not to include cerebral phenomena, like all other phenomena of living bodies, in the laws of scientific determinism."[5]

A more direct progenitor of connectionist AI is the psychologist Edward Lee Thorndike. Thorndike who in his quest for behavioral atoms asserted the following:

> If I attempt to analyze a man's entire mind, I find connections of varying strength between a) situations, elements of situations, and compounds of

> situations and b) responses, readiness to respond, facilitations, inhibitions, and direction of responses. If all these could be completely inventoried, telling what the man would think and do and what would satisfy and annoy him in every conceivable situation, it seems to me that nothing would be left out...*Learning is connecting. The mind is man's connection system.*[6]

In order to better appreciate Thorndike's notion of connectionism, a brief and elementary discussion of the nature and activity of neurons will be helpful.

The human central nervous system is composed of more than 100 billion neurons. Neurons together can be considered the functional units of the nervous system, with each neuron consisting of a cell body, axon and dendrites. An axon is the part of the neuron that transmits impulses away from the cell body. Dendrites are the branching processes of a neuron that conduct impulses toward the cell body. The input signal enters the cell almost entirely through the presence of synapses on the neuronal dendrites or cell body. On the other hand, the output signal travels through an axon giving off many separate branches to other parts of the brain, the spinal cord, or the peripheral body. An axon functions like a telegraph wire taking the message from one neuron to another.

The major function of the nervous system is to "process incoming information in such a way that appropriate motor responses occur."[7] The brain considers the vast majority of sensory information to be irrelevant – for the most part, one pays little attention to the parts of the body that are in constant contact with clothing, the pressure involved while sitting among other things. Once the irrelevant sensory information has been filtered out, the important sensory information is directed into proper motor regions of the brain to cause the desired output or response. The directing or channeling of information is called the "integrative function of the nervous system."[8]

Signal transmission plays an important role in information processing in the brain. The synapses that occur at the junction between one neuron and another primarily determine the direction of the nerve signals spread throughout the nervous system. There are both facilitatory and inhibitory signals that control the nature of synaptic activity, either facilitating or inhibiting synaptic transmission. The bottom-line is that information transmission and processing simply takes place through a connection system.

Based on his studies of the mind's 'connection system' in terms of stimuli and responses, Thorndike suggests that one could formulate a theory of how learning takes place. If learning is largely a result of connections between stimuli and responses, then the connection either exists or does not exist. If it exists, it exists at a strength that is either reinforced or diminished as a function of many different variables. S-R models inspired by Pavlov and other behaviorist psychologists insist that learning is not an accretion of passive facts, but a result of active responses that take place in a dynamic environment based on trial and error.

Under this model, there was no use for mentalistic concepts such as consciousness or introspection that could not be classified as variables, and hence were not empirically verifiable. The salience of this model is that nothing is taken for granted; all action should be studied in terms of stimuli and response. For instance, if an animal is hungry in situation S, and if some response R is immediately followed by food, then the habit of responding R in some situation S is automatically reinforced. Therefore learning was largely construed in terms of the strength of an association between a stimulus and response. Based on S-R models, Thorndike proposed the laws of exercise and effect. The laws of exercise simply states that if a given situation is more often followed by a particular response, the associative bond between them will be particularly stronger (Thorndike himself later on revised his views on the laws of exercise, by stating that whatever works with cats does not necessarily work for humans – repetition for humans does not always guarantee learning). The law of effect states that if a response produces satisfying effects, it will most likely be repeated when the situation arises again. Thorndike's formulations were based on empirical evidence inferred from his experiments with cats.

The experiment is as follows – hungry cats are placed in a closed cage where they could see food placed right outside the cages. The cage door opens by hitting a pole that is placed inside. At first, the cats clawed and banged the sides of the cage. During the process, the cat hit the pole and opened the door. These experiments were repeated again, and the cat eventually figured out through trial and error that hitting the pole is the key. Thus, learning takes place through trial and error, and the stronger the learning gets when the associative bond between the stimulus and the desired response is stronger.

In Thorndike's words:

> If a *S-R* connection has a satisfying after-effect which causes some control in the *N* to send forth a confirming reaction, and if the *S* continues, the confirming reaction tends to cause a continuance or continued repetition of the *R* then and there, and often with more vigor and shorter latency. If the situation has

> vanished, the strengthening of the *C* can only manifest itself when *S* recurs, which may be in a few seconds or only after months. There will then be an increased probability of repetition over what there would have been if no confirming reaction had affected the *C* in question. In either case the strengthening causes the repetition, not the repetition the strengthening.[9]

Where *N* stands for the neurons, *C* stands for any activity, state or condition of *N*, *S* stands for any situation or state of affairs considered as a cause of some *C* and *R* stands for any response or state of affairs.

Thorndike performed numerous other experiments where the greater the connection strengths between *S* and *R*, the more predictable the response. Thorndike performed verbal experiments (on the pleasantness of certain English words to English-speaking adults) to ascertain his hypothesis. Over 1,000 words were picked and ranked by 64 educated adults on a scale of 1 to 10, with 1 being the least desirable and 10 being very pleasant. 800 of these words were arranged in 20 sets of 40 words each, 4 with a value 10, 4 with a value 9, 4 with a value of 8 and so forth. The 40 words were arranged in random order. These ranked words were used in an experiment used by a different set of 18 educated adults. When a word was said out aloud, the subjects were asked to pick a number ranging from 1 to 10 to describe the pleasantness or unpleasantness of a word. The experimenter would say "Right," "Close," "Wrong" or "no-announcement" depending on the comparison of the subject's value with the pre-selected values of the 64 judges. During the course of the experiment in which 5 trials were allowed, Thorndike observed that when particular stimulus-response sequences were followed by pleasure, these responses tend to be reinforced or stamped in, while responses followed by unpleasantness or pain tend to be stamped out.

Behaviorists construed that such experiments establishing connection strengths between stimuli and responses painted a more accurate picture of learning, than the models of learning posited by pre-behavioristic psychology. Some of the auxiliary laws formulated by Thorndike include 'multiple responses' (in any given situation, the organism will respond in a variety of ways if its first response is not immediately satisfactory), 'set or attitude' (deals with the predisposition of acting in certain ways – for instance, there might be a tendency to act aggressively towards aggressive behavior), 'prepotency of elements' (learners only react to important aspects or elements of a problem and ignore irrelevant aspects in learning), 'response by analogy' (a person learns in new situations by the resemblance it may have to a prior experience) and 'associative shifting' (this law states that organisms may use similar responses from one stimulus to another. Associative shifting is similar to a stimulus substitution theory where one stimulus may come to represent another stimulus, eliciting the same response from an organism. For instance, Pavlov's dog starts salivating at the sound of the bell even when no food is placed after it gets conditioned to think that the sound of the bell and food come together). All these experiments underscored an important shift from mentalistic abstractions to a more concrete and empirical mode of investigating the mind.

In Thorndike's words:

> A generation ago the study of the human mind had described powers and processes such as abstraction, generalization, judgment, reasoning, choice,

> desire, motives, purposes, attention, the association of ideas, fusion, habit and automatization; but it had made little progress in working out the dynamics by which these functions operate to make men think and act and learn. Since then, experiments with human subjects have led to simpler and more fruitful accounts of the causation of human behavior, but the simplest and perhaps the most fruitful have come from the study of animal learning. The formation of mental connections by varied reaction and the strengthening of one of the variants (the so-called animal learning by "trial and error" or "trial and success") has been found at the base of much of human learning. So also has the process of associative shifting whereby a response first connected with a total, say *ABCDE*, is later aroused by *BCDE*, or *CDE*, or *DE*, or *E* alone. It is not important to decide whether more gain has come from using simple general principles found in animal learning to explain the complex and subtle forms of human learning than from studying the latter alone. The former has surely been useful.[10]

Overall, the extrapolation of findings of simple experiments with animals and humans to eventually explain all other mental phenomena gained rhetorical appeal among psychologists, neuroscientists among others. On a larger scale, these experiments imply that much of the scenic scope of human mental experience usually described in a mentalistic vocabulary could either be explained away or operationalized in terms of a discrete vocabulary of empirically verifiable terms, based on observations from animal behavior. While Thorndike's findings were based on observable behavior, the cause for this observable behavior was ascribed to neuronal interaction in the brain. Early work in connectionist psychology seem to have inspired the advent of connectionist AI, although the indebtedness to Thorndike or other behaviorists is rarely mentioned. Part of the reason is that current models have substantially improved upon earlier S-R models. Second, computers are largely taking the place of animals since computational models, whether symbolic or biologistic, are considered closer cousins to human cognition. This is so because computers can simulate verbal, human behavior unlike animals; except for a few resourceful alphabet-learning chimpanzees.

Neural Modeling to Connectionist AI

Connectionism represents a departure from symbolic AI, yet its antecedents, namely some of the earlier work in neural nets did not abruptly sever the umbilical cord of symbolic AI. The shift from symbolic AI to connectionism started in a gradual and almost imperceptible manner, before it became significantly obvious. In effect, tracing the pathway from early neural modeling to connectionist AI and its parting of ways with symbolic AI is central to a rhetorical understanding of the nature of thought presented by connectionism. These rhetorical markers will help foster a contextualized, epistemological understanding of a school of thought that replaces symbol-manipulation with neural nets. Making sense of a vast body of literature is not always easy, yet in a modest fashion this section will highlight at least the important contours to chart a rhetorical topography of connectionism.

As early as 1938, Claude Shannon postulated the equivalence of the on or off processes of electronic switches with the flow of information – in other words, a transmission model of communication was posited with the primary emphasis on information and not on meaning, since strictly from an engineering point of view semantic content is purportedly irrelevant while talking about the relaying of electrical signals.[11] Electrical signal processing in computers is dependent on a basic unit - a bit, the abbreviated form for b(inary) (dig)it. A bit refers to the amount of information required in the selection of one electrical signal or impulse between two alternatives.

A bit works in the binary sense of 'either-or' logic. If there are four equally probable alternatives, two bits are required to decide on a course of action. Therefore electrical information processing works in the form of reducing electrical impulses by either-or halves.[12]

In their influential piece entitled "A logical calculus of nervous activity," Warren S. McCulloch and Walter H. Pitts theorized that the all-or-none firing properties of neurons are comparable to the binary fashion of electronic switching. They also declared that neural nets in the brain were equivalent to Turing machines at some functional level of computation (implying that studying the microstructure or the neural hardware of the brain, alongside the activities within the nervous system unlocked the key to understanding and replicating mental phenomena):[13]

> Every net, if furnished with a tape, scanners connected to afferents, and suitable afferents to perform the necessary motor-operations, can compute only such numbers as can a Turing machine...each of the latter numbers can be computed by such a net; and that nets with circles can be computed by such a net; and that nets with circles can compute, without scanners and a tape, some of the numbers the machine can, but no others, and not all of them. This is of interest as affording a psychological justification of the Turing definition of computability and its equivalents...[14]

Before arriving at the conclusion that Turing machines and neural nets were equivalent, McCulloch and Pitts made five assumptions in order to facilitate the process of neural representation via symbolic logic:

> The activity of the neuron is an 'all or none' process.
> A certain fixed number of synapses must be excited within the period of latent addition in order to excite a neuron at any time, and this number is independent of previous activity and the position of the neuron.
> The only significant delay within the nervous system is synaptic delay.
> The activity of any inhibitory synapse absolutely prevents excitation of the neuron at that time.

The structure of the net does not change with time.[15]

Since nervous activity is either inhibited or facilitated depending on the synapses, the all or nothing law of nervous activity enables formalizability through logical propositions. The physiological relations existing between nervous activities are construed in such a manner implying some level of correspondence between the propositions and neuronal activity. In other words, every reaction of a neuron is represented by a simple proposition. Having made these assumptions, McCulloch and Pitts represented the neurons, the neurons of a given net, the state of excitation, time and other variables by working out a model, that provides a logical calculus of nervous activity in the brain at a given time or a given states. Logical propositions were employed as a means of an internal representation of the net. Formal and not so much factual equivalences are asserted, so that "many formal neurons might be needed to embody a single property of a real living neuron."[16] McCulloch and Pitts' project revolved around a quest for a direct representation of what they called knowledge; neural processing in terms of logical propositions. However, McCulloch and Pitts fully realized that the equivalences were of a formal, and not so much a factual nature.

In spite of formulating an elegant system, McCulloch directed the attention to developing an effective calculus for intentional relations, implying that computational knowledge should both have the "descriptive power and semantic robustness similar to that of natural languages."[17] J. Mira in his assessment of McCulloch's corpus has the following to say:

> W. S. McCulloch's entire body of work can be considered a search for the representation of knowledge directly at the level of neural processors

> operationally associated with propositional logic. He looked for the neurophysiological epistemology, that is to say, the direct representation of high-level functions in networks of neurons while nonetheless acknowledging that even if we were familiar with the functioning of every single neuron there would still be something missing in our understanding of the global function Nervous Systems (NS). Relational structures, evolutive history, culture, sociological factors and all the external knowledge, always injected in the reduction and interpretation of anatomo-physiological processes at the symbolic level in the domain of the external observer, are necessary to understand the embodiments of mind.[18]

The highlight of McCulloch's contribution is his cybernetic motivation to understand the brain as a digital system. As Mira explains, "the brains are neurophysiological implementations of mind, the problem in neuroscience is one of reverse engineering. Given a set of brains (nervous systems) in the phylogenetic* and ontogenetic* context," the important task at hand is to "specify input signals...to find the set of specifications from which it originates."[19] McCulloch attempted to specify the neurophysiological conditions of the brain as a computational process, and thereby addressed the epistemological question of internal representation by speaking of the mind as connection system that processes information through the input-output model. Paying attention to this issue of representation of knowledge, would in McCulloch's mind address the question of why the mind is in the head:

> Machine evolution demands to know why the mind is in the head...we cannot afford to carry out any computations, no matter how simple, in a hundred parallel paths and demand coincidence. Consequently, no computing machine is as likely to go right under conditions as various as those we undergo...*why is the mind in the head? Because there, and only there, are hosts of possible connections to be formed as time and circumstance demand.*[20]

McCulloch proposed a layered computational model of neural activity with "cooperative modules that sample information in input and output spaces."[21] None less than John Von Neumann credits McCulloch's conception of neural modeling as playing an instrumental role in the development of digital computers, since the two-valued propositional logic of neural circuitry made the assertion of equivalence between neurons and digital elements possible at some level of abstraction (it is worth quoting him at length):

> Every digital computing device contains relay-like elements, with discrete equilibria. Such an element has two or more distinct states in which it can exist indefinitely...In existing digital computing devices various mechanical or electrical devices have been used as elements: (wheels, telegraph relays) and tubes...It is worth mentioning, that the neurons of the higher animals are definitely elements in the above sense...Following Pitts and McCulloch we ignore the complicated aspects of neuron functioning...It is easily seen, that these simplified neuron functions can be imitated by telegraph relays or by vacuum tubes....Since these tube arrangements are to handle numbers by means of their digits, it is natural to use a system of arithmetic in which the digits are also two-valued. This suggests the use of the binary system. The analogs of human neurons, discussed are equally all-or-none elements. It will appear that they are quite useful for all preliminary, orienting considerations on vacuum tube systems. It is therefore satisfactory that here too, the natural arithmetical system to handle is the binary one.[22]

The legacy of McCulloch and his collaborators can be extended beyond the serial computation of von Neumann machines. For instance, Pitts and McCulloch's influential paper "How we know Universals," outlines a theoretical construction of neural networks engaged in pattern recognition, by hypothesizing how visual input could control motor output through the distributed activity of a layered neural network without the need for an executive control.[23] In fact, this piece by Pitts and McCulloch that dispensed with the need for an executive control showcases one of the earliest forays into connectionism.

Artificial neural networks received further confirmation from the publication of a classic essay "What the Frog's Eye tells the Frog's Brain" reinforcing Pitts and McCulloch's initial impressions – namely, that a) an important part of processing information in the brain is through the organized activity of distributed layers of neurons, b) a collection of neurons can effectively carry out a computation without seeking recourse to a central executive control – the study added a further finding that c) the retina begins the transformation of the visual input information that is necessary for the accomplishment of a particular action in the organism (frogs need food and to evade predators regardless how bright or dim the surroundings are).[24] The authors conclude that "by transforming the image from a space of simple discrete points to a congruent space where each equivalent point is described by the intersection of particular qualities in its neighborhood" one could account for the image in "terms of distributions of combinations of those qualities. In short, every point is seen in definite contexts."[25]

However, it must be mentioned that McCulloch and his collaborators refrained from *tout court* conflations of the distinction between brains and computers; instead they found it plausible to assert equivalences between biological and mechanical systems, thanks largely to a parallel and converging emergence of the study of cybernetics of which McCulloch eventually became a key player.[26] Cybernetics, as is well known, pertains to the study of systems of control and communication in animals and machines in terms of feedback mechanisms. Starting as the brainchild of Norbert Wiener, cybernetics initially encompassed the fields of mathematics, neurophysiology and control engineering; but later on expanded to include mathematical logic, psychology, socioeconomics and automaton theory.[27] Cybernetics attempts to show that self-organizing systems are explainable on the basis of principles such as feedback (positive and negative), homeostasis, entropy, information (transmission of messages through a communication channel) and negentropy (negative entropy). From a rhetorical point of view, the importance of cybernetics is that it engenders a cross-fertilization of ideas between disciplines studying biological and mechanical systems respectively, making one explainable in terms of the other and vice-versa. Notions such as genetic 'codes,' mental 'programs' only to name a few are exemplars of such confluence that enables the blurring of boundaries between biological and mechanical, real and artificial and living and non-living things. Norbert Wiener's reflections on his conversation with Pitts is insightful in highlighting the emerging nexus between neurology and engineering:

> At that time, Mr. Pitts was already thoroughly acquainted with mathematical logic and neurophysiology, but had not had the chance to make very many engineering contacts. In particular, he was not acquainted with Dr. Shannon's work, and he had not had much experience with the possibilities of electronics. He was very much interested when I showed him examples of modern vacuum tubes and explained to him that these were ideal means for realizing in the metal the equivalents of his neuronic circuits and systems. From that time, it became clear to us that the ultra-rapid computing machine, depending as it does on consecutive switching devices, must represent an almost ideal model of the problems arising in the nervous system. The all-or-none character of the discharge of the neurons is precisely analogous to the single choice made in determining a digit on the binary scale, which more than one of us had already contemplated as the most satisfactory basis of computing-machine design.[28]

The cybernetic movement to equate the biological and the mechanical had its own internal fissures as well. Wiener himself disavowed the determinist materialism of classical physics and acknowledged the role of indeterminacy in biological systems. K. M. Sayre reflecting on Wiener's position states, "a theoretical basis for rejecting determinism in the natural world generally lies in the principle that all irreversible processes tend to involve a loss of negentropy, which entails that causes generally tend to be more highly structured than their effects."[29] Wiener comments that even the most "complete collection of data for the present and the past is not sufficient to predict the future more than statistically."[30] Eventually cyberneticians distanced themselves from the trajectory of Laplacian determinism that symbolic AI was pursuing, thus gradually uncoupling ties between the two:

> Cybernetics has been guided from the outset by the conviction that a wide range of human mental functions can be reproduced mechanically. Among Wiener's original associates in the 1940s were several figures (e.g. O. G. Selfridge, W. H. Pitts, W. S. McCulloch) who subsequently became known for contributions to AI. Spokesmen for cybernetics up through the late 1970s, (e.g. F. H. George, K. Gunderson, K. M. Sayre) still considered AI to be an integral part of the discipline. The original ties between cybernetics and AI were effectively severed during the 1980s, however to the extent that early contributors to machine intelligence who had remained closely identified with the former movement (e.g. W. R. Ashby, D. M. MacKay, F. H. George, Wiener himself) are seldom cited in current histories of the latter. The divorce appears to have been due largely to the takeover of AI by the computational paradigm, and to an ideological slide towards materialism on the part of its advocates.[31]

Symbolic AI became the dominant version of AI, and as discussed in the previous chapter it follows a top-down approach beginning with the assumption that humans are machines, and that computational models of mind are adequate representations of neural activity. Even early work in neural nets focused on the merited sufficiency of mechanical models of mind. McCulloch and Pitts' seminal piece on neural nets rests on the assumption that logical nets are exemplars for neural nets in the brain, although the authors acknowledge formal equivalences instead of factual equivalences (the link was made by arguing that the all-or-nothing laws of nervous activity coincided with the binary logic of computational devices) – even so, with an eye on preempting criticism declare that the formal equivalences of neural nets are by themselves sufficient indices of mental activity and provide a comprehensive theory of nervous activity:[32]

> The importance of formal equivalence lies in this: that the alterations actually underlying facilitation, extinction, and learning in no way affect the conclusions which follows the formal treatment of the activity of nervous nets, and the relation of the corresponding propositions remain those of the logic of propositions.[33]

In short, it was customary to invoke the sufficiency of logico-mathematical models either while discussing the nature of mental activity or while attempting to construct machines that could simulate mental activity.

Against this backdrop, F. Rosenblatt's probabilistic model for information storage and organization in the brain marks the beginning of the departure from symbolic AI, and probably the earliest foray into connectionism:

> During the last few decades, the development of symbolic logic, digital computers, and switching theory has impressed many theorists with the functional similarity between a neuron and the simple on-off units of which computers are constructed, and has provided the analytical methods necessary for representing highly complex logical functions in terms of such elements. The result has been a profusion of brain models which amount simply to logical contrivances for performing, particular algorithms (representing "recall," stimulus comparison, transformation and kinds of analysis) in response to sequence of stimuli…*Unfortunately, the language of symbolic logic and Boolean algebra is less well suited for such investigations*. The need for a suitable language for mathematical analysis of events in systems where only the gross organization can be characterized, and the precise structure is unknown, has led the author to formulate the current model in terms of probabilistic theory rather than symbolic logic. *The theorists (referring to the*

> *symbolists) referred to above were chiefly concerned with the question of how such functions as perception and recall might be achieved by a deterministic physical system of any sort, rather than how this is actually done by the brain.*[34]

Rosenblatt's call for a shift from top-down mathematical models to bottom-up biological models is a renewal and strikingly reminiscent of Bernard's postulation that deterministic models from mathematics are inadequate to study living systems. Rosenblatt puts forward three fundamental questions: "1) How is information about the physical world sensed, or detected, by the biological system? 2) In what information is stored, or remembered? 3) How does information contained in storage, or in memory, influence recognition and behavior?"[35] To answer these questions, Rosenblatt posits his perceptron as a probabilistic model – by suggesting that his model is based on the following assumptions gleaned from theorists such as Hebb, Hayek, Uttley and Ashby who were more directly concerned with the biological nervous system:

> The physical connections of the nervous system which are involved in learning and recognition are not identical from one organism to another. At birth, the construction of the most important networks is largely random, subject to a minimum number of genetic constraints.
> The original system of connected cell is capable of a certain amount of plasticity; after a period of neural activity, the probability that a stimulus applied to one set of cells will cause a response in some other set of cells will cause a response in some other set is likely to change, due to some relatively long-lasting changes in the neurons themselves.
> Through exposure to a large sample of stimuli, those which are most "similar" (in some sense which must be defined in terms of the particular physical system) will tend to form pathways to the same sets of corresponding cells. Those which are markedly "dissimilar" will tend to develop connections to different sets of responding cells.
> The application of positive and/or negative reinforcement (of stimuli which serve this function) may facilitate or hinder whatever formation of connections is currently in progress.

> Similarity, in such a system, is represented at some level of the nervous system by a tendency of similar stimuli to activate the same sets of cells…The structure of the system, as well as the ecology of the stimulus-environment, will affect, and will largely determine, the classes of "things" into which the perceptual world is divided.[36]

Based on the preceding assumptions, Rosenblatt hypothesized the architecture of the perceptron (that would engage in pattern recognition among other things) according to the following rules: 1) according to the all or nothing law of nervous activity, a set of stimuli are said to impinge on a retina of sensory units (S-points), 2) the impulses from contact are channeled to a set of cells, known as association cells (A-units) in a projection area – the cells in the projection area individually receive a number of connections from the sensory points. The sensory units transmitting impulses to a particular association cell (A-unit) are referred to as origin points. These origin points may be either inhibitory or excitatory. The origin points of A-units are distributed through the network, 3) from this interaction, there are sets of emergent response units with either excitatory or inhibitory feedback depending on the strength of the connections between the sensory units and the association cells. Rosenblatt concluded that "the fundamental phenomena of learning, perceptual discrimination, and generalization can be predicted from six physical parameters," namely the number of excitatory connections for each association cell, the number of inhibitory connections for each association cell, the expected threshold, the proportion of response units to each association cell, the number of association cells and the number of response units.[37]

All this just means that the brain is a parallel-processing computer, instead of a serial-processing von Neumann machine. Paying adequate attention to the neural pathways and neural interaction provides modelers with a biological basis for mental cognition, that computationalists simply ignored or failed to provide. In rhetorical terms, mathematics has hitherto represented the perfect language, unadulterated by the semantic quirks of every day use – but the shift from an almost perfect, but static conception of knowledge to a more dynamic and organic but indeterminate model borrowed from biology and statistics signifies a move to experiment with, if not embrace contingency. While there can be no impugning of the sense of order in the structural makeup of organization, the effects or the way in which living systems operate, for instance say behavior, is not at least apparently deterministic. The lure of mathematical logic revolves around the principles of orderliness and consistency. On the other hand, biologistic models have the rhetorical appeal of bearing veritistic semblances to organic systems.

Even so, Rosenblatt's perceptrons, despite its organic promise, ran afoul with the symbolic AI community, as Marvin Minsky and Seymour Papert recall:

> Rosenblatt's schemes quickly took root, and soon there were perhaps as many as a hundred groups, large and small, experimenting with the model...the results of these hundreds of projects and experiments were generally disappointing, and the explanations inconclusive. The machines usually work quite well on very simple problems but deteriorate very rapidly as the tasks assigned to them get harder.[38]

Regardless of the negative indictment of Rosenblatt's project, the salience of his approach vis-à-vis symbolic AI is best stated in his own words:

> The implicit assumption (of the symbolic program) is that it is relatively easy to specify the behavior that we want the system to perform, and that the challenge is then to design a device or mechanism which will effectively carry out this behavior...*It is both easier and more profitable to axiomatize the physical system and then investigate this system analytically to determine its behavior, than to axiomatize the behavior and then design a physical system by techniques of logical synthesis.*[39]

Although it was not uncommon for pioneers in both the schools of thought to overstate the scope and significance of their respective projects, Rosenblatt's perceptrons were, according to David Rumelhart and James McClelland unmistakably precursors of PDP processing, that came to define much of the impulse of connectionism:

> Rosenblatt's work was very controversial at the time, and the specific models he proposed were not up to all the hopes he had for them. *But his vision of the human information processing system as a dynamic, interactive, self-organizing system lies at the core of the PDP approach*...The studies of perceptrons...clearly anticipated many of the results in use today. The critique of perceptrons by Minsky and Papert was widely misinterpreted as destroying their credibility, whereas the work simply showed limitations on the power of the most limited class of perceptron-like mechanisms, and said nothing about more powerful, multiple layer models.[40]

The emergence of PDP networks to redress the lacunae in symbolic AI that was traditionally good at building expert systems but inefficient at performing tasks such as pattern recognition among others, rehabilitated perceptrons by giving it at least favorable mention in histories of connectionist AI. Before discussing about the elements of connectionism with respect to PDP, it is instructive to note that none less than Minsky himself (regarded as one of the chief morticians for Rosenblatt's perceptrons) speaks at least accommodatingly of connectionism in his assessment of the field of AI as a whole:

> Which approach is best to pursue? The answer is simple: we have to use both. In favor of the top-down side, research in AI has told us a little – but only a little – about how to make machines solve problems by using methods only a little – about how to make machines solve problems by using methods that resemble reasoning. In favor of the bottom-up approach, the brain sciences told us a little – but again, only a little – about what brain cells do. If we knew enough more about brain cells and their connections we could try to work from that to discover how they support our higher level processes. If we understood more about thinking we work down toward finding out how brain cells do it. But right now we're caught in the middle; we know too little at either extreme. The only practical option is to ping-pong between them, searching for materials with which to build a plausible bridge. How can we do that? One way is to focus on inventing various ways to represent knowledge, and then to try to extend those techniques in both directions. On the connectionist side we can try to design neural networks that can learn those representations. Then we can try, on the top-down side, to design higher level systems that can effectively exploit the knowledge thus represented.[41]

Rhetorically speaking, trying to integrate a top-down, synthetic approach with a bottom-up, biological approach to computation may require coming up with an intermediate metaphor or model of mind. In some ways, that might muddle the existing models of cognition – what that would like remains to be seen.

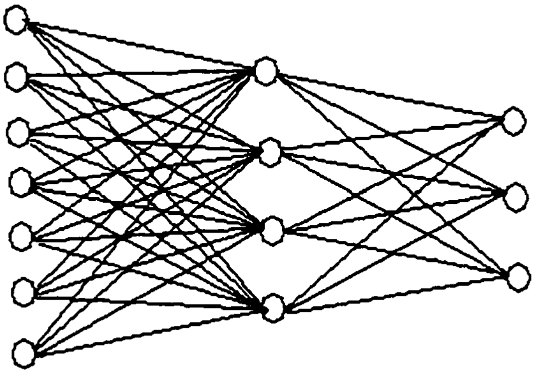
Parallel Distributed Processing

As stated in the previous chapter, Symbolic modeling ran aground while attempting to take on tasks that require real-time, real-world interaction. The simpler tasks that humans normally take for granted such as visual and pattern recognition among others were not realizable in symbol systems. More importantly, as Paul Smolensky states, "insights into the design and implementation of physical symbol systems have so far shed virtually no light on how the brain works."[42] Connectionist AI, on the other hand, premised on the promise that biological modeling is the answer suggests that large networks of interconnected and simple computational units effected through parallel processing would widely enhance the scope of Artificial Intelligence. Each computational unit has an activation value in the form of a number, which is communicated to other connected units or processors of varying strengths. The activity of the units or processors directly influences the changes in the activation value. For instance, if the activation values of some of the units form the input and others the output, the connection between the units is instrumental in determining how the input is transformed into output. As Smolensky explains, "in connectionist systems, knowledge is encoded not in symbolic structures but rather in the pattern of numerical connections between units."[43]

The explicit assumption is that the interworkings of the neurons from the central nervous system provide a good model for artificial neural networks. The entire cerebral cortex could be considered as a large neuronal pool, or a collection of smaller pools. There are several neuronal pools that include the basal ganglia, the specific nuclei in the thalamus, in the cerebellum, the mesencephalon, pons and medulla. Each pool possesses distinctive characteristics pertaining to its organization that causes signal processing in its own special way, enabling the realization of a "multitude of functions of the nervous system."[44] The relaying of signals take place through neuron pools – each input fiber divides ranging from a few hundreds to thousands of times, branching out into a large number of terminal fibrils that spread over a "large area in the pool to synapse with the dendrites or cell bodies of the neurons in the pool."[45] The individual neurons are acted upon by large numbers of fibers in order to cause an excitatory stimulus, and these neurons are considered facilitated. Incoming fibers can also inhibit neurons, rather than exciting them. Guyton illustrates this point well, as follows,

> Sometimes an incoming signal to a neuronal pool causes an output excitatory signal going in one direction and at the same time an inhibitory signal going elsewhere. For instance, at the same time that an excitatory signal is transmitted by one set of neurons in the spinal cord to cause forward movement of a leg, an inhibitory signal is transmitted simultaneously through a separate set of neurons to inhibit the muscles on the back of the leg so that they will not oppose the forward movement.[46]

As David Rumelhart and James McClelland, the model of a brain as a parallel processing system is the driving impetus behind Parallel Distributed Processing models: "these models assume that information processing takes place through the interactions of a large number of simple processing elements called *units*, each sending excitatory and inhibitory to other units."[47] The arborization process of neurons in the brain serves as the original prototype for neural networks. Paul M. Churchland describes the operations of artificial neural networks as simulations of natural neurons, through artificial units that admit various levels of activation values between 0 and 1.[48] And the level of activation is primarily a function of the "number of connections, of their size or weight, of their polarity (stimulatory or inhibitory) and of the strength of incoming signals."[49]



For instance, the diagrammatic representation of a very simple neural net, shows at least three layers of units. Each input unit or processor is said to have an activation value that represents either some environmental stimuli or feature external to the net. The activation values of each input units are send to the each of the hidden units to which they are connected. The corresponding hidden units respectively calculate its own activation value depending on the activation value of the input units. Signals are send to the output units that comes up with its own activation values based on the preceding process of activation at the hidden and input units.

The entire process or pattern of activation is dependent on the structure of the net, in turn determined by the weights or strengths of connections between the units. The weights are seen as negative or positive, with a negative weight representing the inhibition of the signal at the receiving unit, and a positive weight representing the facilitation of the signal at the receiving unit. The activation values at the output units are calculated by a simple activation function, the function is said to sum together the contributions of all transmitting units, where the contribution of a unit is defined as the weight of the connection between the sending and receiving units times the sending unit's activation value. Further, the sum could be altered by adjusting the activation sum to a value between 0 and 1 and/or by setting the activation to a zero unless a threshold sum could be reached. Connectionists stake their claims on the principle that all of cognitive activity can be explained as an interaction between interconnected units, human cognition can be explained by the studying the strengths of connections between units. It must be noted here that the key difference between symbolic and connectionist AI, is that the latter dispenses with the notion of a central processing unit:

> There is no central processor or controller, and also no separate memory or storage mechanism. The only activity in the system is these little units changing state, in response to signals coming in along these connections, and then sending out signals of their own. There are two ways in which such a network can achieve a kind of memory. First in the short term, information can be retained in the system over time insofar as the units tend to change state only slowly (and, perhaps, regularly). Second, and in the longer term, there is a kind of memory in the connections themselves. For, each connection always connects the same two units (they don't move around); and, more significant, each connection has a property, called its "weight" or "strength," which is preserved over time.[50]

Connectionists claim not only isomorphism but also a greater veridical similarity between artificial neural networks and organic brains. Connectionism also promises to take the development of AI further than symbolic AI, by addressing the shortcomings of the symbolic paradigm:

> From the perspective of neuroscience, the problem with the symbolic paradigm is quite simply, as I have already indicated, that it has provided precious little insight into the computational organization of the brain. From the perspective of modeling human performance, symbolic models...do a good job at a coarse level; but the finite structure of cognition seems to be more naturally described by nonsymbolic models. In word recognition, for example, it is natural to think about activation levels of perceptual units. In AI, the trouble with the Boolean

> dream is that symbolic rules and the logic used to manipulate them tend to produce rigid and brittle systems.[51]

Smolensky considers the connectionist approach to be the subsymbolic paradigm that serves as an "intermediate level of structure between neural and symbolic levels."[52] The subsymbolic paradigm attempts to formalize, "at some level of abstraction, the kind of processing occurring in the nervous system," and although "many details of neural structure and function are absent from the subsymbolic level, and the level of description is higher than the neural level," it seems "quite clear that connectionist systems are much closer to neural systems than are symbolic systems."[53] The divergence from the rules-based, symbol manipulation approach of traditional AI purportedly leads to a "reconceptualization of key psychological issues, such as the nature of the representation of knowledge," whereby "such knowledge is represented, often in widely distributed form, in the connections among the processing units."[54] The parallel processing and distributed representations among collections of units manifests certain emergent properties such as cognition and so forth. David E. Rumelhart, James L. McClelland and the PDP research group claim that mental categories such as understanding, cognition among other things should be seen as the manifestation of neural activity in parallel processing, therefore models that simulate neural activity should operate on the principle of parallel distributed processing:

> These models assume that information processing takes place through the interactions of a large number of simple processing elements called units, each sending excitatory and inhibitory signals to other units. In some cases, the units stand for possible hypotheses about such things as the letters in a particular sentence. In these cases, the activations stand roughly for the strengths associated with different possible hypotheses, and the interconnections among the units stand for the constraints the system knows to

> exist between the hypotheses. In other cases, the units stand for possible goals and actions, such as the goal of typing a particular letter, or the action of moving the left index finger, and the connections relate goals to subgoals, subgoals to actions, and actions to muscle movements. In still other cases, units stand not for particular hypotheses or goals, but for aspects of these things. Thus a hypothesis about the identity of a word, for example, is itself distributed in the activation of a large number of units.[55]

Parallel distributed processing takes us to unchartered territory much beyond the scope of Rosenblatt's single layer perceptrons. In spite of their "physiological plausibility and neural inspiration," Rumelhart et al state that the primary appeal of PDP models are for "psychological and computational reasons," since they provide hope of offering "computationally sufficient and psychologically accurate mechanistic accounts of the phenomena of human cognition which have eluded successful explication in conventional computational formalisms, " and have even more change the way in which we think about representation and the mechanisms of learning.[56]

The PDP team further asserts that their models offer "alternatives to serial models of the microstructure of cognition," and describes the "internal structure of the larger units, just as subatomic physics describes the internal structure of the atoms that form the constituents of larger units of chemical structure."[57] The significant difference between PDP models and other models is that the notion of representation is significantly altered – instead of looking for a single source, code or node as the key for representation, PDP instead contends:

> The representation of the knowledge is set up in such a way that the knowledge necessarily influences the course of processing. Using knowledge in processing is no longer a matter of finding the relevant information in memory and bringing it to bear; it is part and parcel of processing itself. *For learning, the implications are equally profound. For if the knowledge is the strength of the connections, learning must be a matter of finding the right connection*

> *strengths so that the right pattern of activation will be produced under the right circumstances. This is an extremely important property of this class of models, for it opens up the possibility that an information processing mechanism could learn, as a result of tuning its connections, to capture the interdependencies between activations that it is exposed to in the course of processing.*[58]

Instead of explicit rules based information processing, it is precisely the connection strengths "which allow a network of simple units to act as though it knew the rules."[59] The learning mechanism is not attributed to any powerful computational abilities, but knowledge is stored in the strengths of interconnections between units. The seven major components of any connectionist system such as

> a set of processing units,
> a state of activation defined over the processing units,
> an output function for each unit that maps its state of activation into an output,
> a pattern of connectivity among units,
> an activation rule for combining the inputs impinging on a unit with its current state to produce a new level of activation for the unit;
> a learning rule whereby patterns of connectivity are modified by experience;
> and an environment within which the system must operate,[60]

reflect the use of "brain-style computational systems," that "offers not only a hope that we can characterize how brains actually carry out certain information-processing tasks but also solutions to computational problems that seem difficult to solve in more traditional computational frameworks."[61]

A few examples of connectionist models that lend credence to the belief that artificial neural nets are good indicators of human cognition are T. Sejnowski and C. Rosenberg's NETtalk, Rumelhart and McClelland's past-tense predictor and J. L. Elman's grammatical nets. Sejnowski and Rosenberg's NETtalk, as their transparent title indicates, is a parallel network that can learn to pronounce English from a fairly large transcript of words.[62] The training for NETtalk was from a large database that consisted of English written text with a corresponding set of phonetic outputs and a speech synthesizer. After a series of trials and errors, NETtalk evolved from producing random noise as its first output, through garbled speech and finally to a state of making competent pronunciations. David E. Rumelhart and James L. McClelland's net learned to predict the past tense of English verbs.[63] The past tense of English verbs can be really confusing for non-native speakers since some past tense words are formed simply by adding the suffix 'ed,' whereas irregular past tenses take form of was, came, went and so forth. The net was first trained with a large number of irregular verbs and later with a set of regular 'ed' ending words. At first, the net had a tendency to overregularize by coming up with words such as 'broked' instead of broken. However, with more training the net self-corrected and rightly predicted the past tense verbs, even going so far as to recognize regularities among irregular verbs (blow/blew, build/built, fly/flew and so on and so forth). An interesting fact pointed out by connectionists is that the net's initial mistakes were comparable to the common grammatical errors among children who commit similar errors of mixing up regular and irregular verbs.

However, without taking any credit from Rumelhart and McClelland's net, it must be pointed out that there are certain linguistic blindspots as well. A case in point is the issue of machine translation – for instance, there is an increasing reliance on the part of Bible translators on software to speed up the process of translating scriptures into the vernacular. On one hand, computer software has significantly reduced the amount of time required to accomplish the task of translation. Yet, many passages may read rather irreverently if there is no human intervention, especially while addressing cultural contexts, idiomatic expressions and culture specific humor among other things that are clearly environmental. Last but not the least, existing neural nets do not, in my opinion, explain how writers coin neologisms or invent new words. For instance, Salman Rushdie invented a word 'chutnification' to refer to the linguistic habits of English speakers in the Indian sub-continent to create intercultural portmanteau words to describe cultural experiences for which the existing vocabularies of neither English nor any one of the many native languages nor dialects would suffice. Interestingly enough, the suffix 'fication' is borrowed from English, whereas chutney is a Hindi word for a sauce made of different spices and mixes. In effect, the invention of new words is a rhetorical reflection of the need to verbalize an experience, when the existing vocabulary is inadequate to account for the context-specific, cultural experiences. Furthermore, the importation of a totally different meaning into a word, the justifications of its invention among other semantic imponderables are factors that make existing neural nets inadequate to account for human communication. Needless to say, none of these criticisms are meant to diminish Rumelhart and McClelland's neural net or the future direction of connectionist nets, but it is important to cover much more linguistic and environmental territory before generalizing about the range of verbal behavior.

Last but not the least, J. L. Elman's development of grammatical nets with an intent on achieving mastery of the rules of grammar, performed efficiently in predicting the next word in a large body of English sentences.[64] From a simple vocabulary of 23 words, sentences were formed using a subset of English grammar. The task although seemingly simple, presented certain linguistic problems in that a large number of relative clauses were allowed while demanding agreement between the principal noun and the verb. In the sentence, 'Any man that chases dogs that chases cats…runs,' the principal noun "man" must agree with the verb "runs," despite the presence of other nouns such as 'dogs' and 'cats' in the middle. The net succeeded in predicting words such as these by following certain grammatical rules. For a statement such as the one mentioned above, the output units for words that are grammatical continuations of the sentence should has to be in the active voice. Elman's nets were able to fill in the gaps, by accurately predicting the ensuing term. All the three examples mentioned above represent significant efforts in the developmental trajectory of neural nets that are more flexible by following statistical patterns and thus accomplish a larger range of activities than classical symbol-processing.

A widely used learning algorithm used in neural nets is usually referred to as backpropagation.[65] Backpropagation as a learning tool is particularly helpful, for instance if a network undergoes a training period during which a series of inputs are presented, necessitating an output for each of the presentation. When a comparison is made between the produced output and the target output for each presentation, an error signal is produced. The error signal is send or propagated back through the network, requiring the adjustments of weights to produce the desired output. Since the weights are kept constant after training, the system is therefore able to make informed guesses when the inputs are provided, and through a repetition of the process the desired output is produced.

Backpropagation and the lack of a central executive presents in some ways a revolutionary or new way to study cognition. However, empirical evidence is still necessary to reinforce the connectionist claims of the isomorphism between human brains and neural nets. The types of equivalences asserted are still formal, and not necessarily factual. As a desideratum, more empirical research is necessary before *tout court* equivalences are asserted. Haugeland's reflections on the brain-machine isomorphism professed in connectionist circles is particularly salient:

> Obviously, connectionist networks are inspired to some extent by brains and neural networks. The active units are like individual neurons, and the connections among them are like the axons and dendrites along which electro-chemical "pulses" are sent from neuron to neuron. But while this analogy is important, it should not be overstressed. What makes connectionist systems interesting as an approach to AI is not the fact that their structure mimics biology at a certain level of description, but rather what they can do. After all, there are countless other levels of description at which connectionist nets are utterly unbiological; and if some GOFAI (symbolic AI) account turns out to be right about human intelligence, then there will be *some* level of description at which it too accurately models the brain. Connectionist and allied research may someday show that neural networks are the level at which the brain implements psychological structures; but this certainly cannot be assumed at the outset.[66]

Yet, in spite of the formal equivalences (as opposed to factual equivalences) the idea of finding symmetry between brains and artificial neural networks rhetorically validates studies of one in terms of the other. The brain becomes a prototype for artificial neural nets, and by the same token, neural nets could provide inspiration for furthering studies in the information processing involved in certain cognitive tasks. Although such comparisons are only natural, given the context of interaction, some philosophers have used connectionism as an empirical prop for eliminative materialism, and eliminative materialism is in turn used as a philosophical platform for furthering connectionism while simultaneously negating alternative accounts of the mind. Instead of pursuing a conciliatory middle road that acknowledges the contributions of different theories of mind, that provide answers for certain types of questions, the quest for a unified theory of cognition is rhetorically pitting one school of thought against the other. In other words, not unlike the 'all or nothing' principles of neural activity, a select few cognitive scientists are engaged in an all-or-nothing quest for a theory of cognition. One such theory is eliminative materialism.

Connectionism and Eliminative Materialism

Eliminative materialism is a philosophy premised on the notion that folk psychological accounts of mind are false, and also inadequate to account for the mind.[67] Paul Churchland describes the project of eliminative materialism as follows:

> Eliminative materialism is the thesis that our common-sense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced, rather than smoothly reduced, by completed neuroscience. Our mutual understanding and even our introspection may then be reconstituted within the conceptual framework of completed neuroscience, a theory we may expect to be more powerful by far than the common-sense psychology it displaces, and more substantially integrated within physical science generally.[67]

Folk psychology pertains to the common sense framework for mental phenomena as a theory brings about a rather unified approach to "most of the major topics in the philosophy of mind, including the explanation and prediction of behavior, the semantics of mental predicates, action theory, the other-minds problem, the intentionality of mental states and the mind-body problem."[68]

Explanations from folk psychology often make reference to terms such as beliefs, desires, fears, intentions, perceptions, and so forth. Churchland and other eliminativists take serious objection to quotidian conceptions of mental phenomena, and state that neuroscience would explain away both common sense perceptions and the sentential logic of propositional attitudes. The conception of mind, postulated by symbolic AI and sentential logic falls within the purview of folk psychology, in that mental phenomena are explained in terms of the manipulation of linguistic symbols, following a set of discrete rules. Churchland disagrees with this notion and instead suggests that:

> Research into the neural structures that fund the organization and processing of perceptual information reveals that they are capable of administering a great variety of complex tasks, some of them showing a complexity far in excess of that shown by natural language. Natural language, it turns out, exploits only a very elementary portion of the available machinery, the bulk of which serves far more complex activities beyond the ken of the propositional conceptions of FP. The detailed unraveling of what the machinery is and of the capacities it has makes it plain that a form of language far more sophisticated than "natural" language, though decidedly "alien" in its syntactic and semantic structures, could also be learned and used by our innate systems. Such a novel form of communication, it is quickly realized, could raise the efficiency of information exchange between brains...since it would reflect the underlying structure of our cognitive activities in greater detail than does natural language.[69]

Churchland attempts to suggest that there is some type of neurological black box that would systematically explain every mental operation. Neuroscientific accounts would in turn exorcise the spells of common sense perceptions primarily mediated by natural language in a cruder form, and also sentential logic where a mentalist vocabulary is articulated in a more sophisticated form. Churchland is perhaps right that neurological accounts of mind are more reliable than intuitionist or logical accounts of mind; however, it seems to this writer that a dialectical model of mind and matter is inevitable at some level or the other. Although there is a neurological basis for mental experience, the articulation of this mental experience, even a scientific exposition of this experience has to be mediated linguistically. Language indeed has a neurological basis, but even the account of the neurological basis of language should seek recourse to language. Yes, language needs neurons in the brain. Yet the explication of neural activity is impossible without language. As long as human experience is mediated linguistically, the realm of metaphysics, folk psychology, natural language or even symbolic logic cannot be simply eliminated. One could replace one set of mental terms with another; even if the latter is scientifically anchored the impossibility of escaping the realm of natural language is a given.

At the heart of eliminative materialism, is the desire to explain away beliefs, desires among others as the basis for human behavior. Folk psychology largely construes human behavior to be intentional and causal. According to common sense:

> (1) when people see a dog nearby they typically come to believe *that there is a dog nearby,*
> (2) when people believe *that the train will be late if there is snow in the mountains,* and come to believe *that there is snow in the mountains*, they will typically come to believe *that the train will be late*;
> (3) when people who speak English say "There is a cat in the yard," they typically believe *that there is a cat in the yard.*[70]

Ramsey, Stich and Garon explain that these "generalizations of common-sense psychology are couched in terms of the semantic properties of the attitudes...the belief that *p* that a given belief has a given effect or cause. Thus common-sense psychology treats the predicates expressing these semantic properties" as "nomological or law-like generalizations."[71] Instead of propositional attitudes, eliminative materialists suggest that connectionist accounts of cognition are more accurate. Ramsey, Stich and Garon cite Paul Smolensky's overview of connectionism at length to base neural networks as the paradigmatic grounds for cognition:

> Connectionist models are large networks of simple, parallel computing elements, each of which carries a numerical activation value which it computes from neighboring elements in the network, using some simple numerical formula. The network elements or units influence each other's values through connections that carry a numerical strength or weight...In a typical...model, input to the system is provided by imposing activation values on the input units of the network; these numerical values represent some encoding or representation of the input. The activation on the input units propagates along the connections until some set of activation values emerges on the output units; these activation values encode the output the system has computed from the input. In between the input and output units there may be other units, often called *hidden units,* that participate in representing neither the input nor the output. The computation performed by the network in transforming the input pattern of activity to the output pattern depends on the set of connection strengths; these weights are usually regarded as encoding the system's knowledge. In this sense, connection strengths play the role of the program in a conventional computer. Much of the allure of the connectionist approach is that many connectionist networks program themselves, that is, they have autonomous procedures for tuning their weights to eventually perform some specific computation. Such learning procedures often depend on training in which the network is presented with sample input/output pairs from the function it is supposed to compute.[72]

The adjustment of connection weights and its subsequent production of outputs resulting in intelligent behavior has a precedent in the S-R models of early behaviorist models, whereby mental categories were considered untenable for a scientific investigation of the mind. Folk psychological theories rooted in sentential logic and the common sense perceptions of mind put forward by laypersons are scientifically inadequate, and hence eliminative materialism turns to the neurosciences for a neural realism as opposed to intuitionist or logical realism.

Churchland turns to artificial neural networks to provide adequate neurological models of all types of mental activity in the brain. In his words:

> If even small artificial networks can perform (such) sophisticated cognitive tasks...there is no mystery that real networks should do the same or better. What the brain displays in the way of hardware is not radically different from what the models contain, and the differences invite exploration rather than disappointment. The brain is of course very much larger and denser than the models so far constructed. It has many layers rather than just two or three...It plainly commands many spaces of stunning complexity, and many skills in consequence. It stands as a glowing invitation to make our humble models yet more and more realistic in hopes of unlocking the many secrets remaining.[73]

Churchland assumes at the outset without much elaboration that the brain's "hardware" is similar to the models' hardware. From the perspective of neural modeling, there is unmistakably much merit to the connectionist accounts of cognition, however what is unclear is why eliminativists divorce 'language like' structures from neurological accounts of cognition. The human brain is sufficiently complex to perform the type of mental activities performed by both symbolic and connectionist models. Certainly, it is not inconceivable that one may find a neurological basis for some aspects of computation as enunciated by symbolic logic just as much as there is a neurological basis for some if not, all aspects of connectionism. The counter-argument is to say that the human brain works more like a neural net, than a serial computing machine.

Churchland however simply has no place for symbolic AI or alternative models of cognition. "In urging the poverty of sentential epistemologies," Churchland suggests that his primary motivation has been the pattern of failures displayed by the classicist approach whose failures imply that "what is defective…is its fundamental assumption that language-like structures of some kind constitute the basic or most important form of representation in cognitive creatures, and the correlative assumption that cognition consists in the manipulation of those representations by structure-sensitive rules."[74] Churchland's notion of 'language-like' is perhaps fuzzy, in some ways, that it could either refer to language like structures that play an important role in the representation of what we consider as cognition, or it could refer to logical forms such as syntax, propositions and so forth.[75] On one hand, Churchland expressly argues that the propositional attitudes of folk psychology, largely based on functionalist accounts of cognition are radically false and misleading. On the other hand, he takes common sense perceptions such as beliefs, desires and so forth as a whole to task. Furthermore, instead of a sentential epistemology Churchland advances a neuroscientific epistemology that replaces the 'language like' structures of human cognition. Churchland is perhaps right, in asserting that rules-like a priori mentalistic impositions derived from sentential logic may not reflect the actual way the brain works, although most die-hard rationalists would like to think so. However, what is unclear, is that even if plain neurological and neurological interactions alone held the key, how one can escape representing the action of neural structures without resorting to some sort of linguistic categorization, which only implies that one level of linguistic categorization is replaced by another – that only goes to show that even the neurosciences are linguistically mediated, even though language in itself has a neurological basis. Furthermore, Churchland himself should seek recourse to logical reasons to justify the ascendancy of the neurosciences. Although, the arguments may not take the form of a symbolic logic equation, at some level or the other, neuroscientists should themselves seek recourse to informal logic or practical reasoning to justify their arguments, even if there is clear and distinct evidentiary grounds from empirical experimentation. Therefore, the question is, can one really escape 'a priori' impositions even in "bottom-up" neurological approaches to cognition? None of this is meant to say that the logical approach presupposed by symbolic AI is more accurate; just that even a neurological approach should at some level or the other seek logical arguments to justify the 'neural realism' of artificial neural networks.

One could perhaps learn a lesson or two from Jay F. Rosenberg's comments about the indispensable logical need to justify the representation of what is being represented:

> Although we do not have the slightest idea how representations having the logical forms of conditionals and negations might be encoded in the "distributed" way appropriate to connectionist networks...and if they can, there is also no reason to reject out of hand the suggestion that the homeostatic "web of belief" interadjustments among such representational and inferential commitments take the de facto operational form of adjustments of the activation weights of the hidden units of some complex multi-layered system of the connectionist sort. The point, however, is that *this* connectionist story is *essentially* the story, not of an alternative to a sentential epistemology, but of the *implementation* of a sentential epistemology. The "language-like" character of the logically articulated representations thus encoded is not

> adventitious, but necessary for us to be able to understand the operation of such a system in *epistemic* terms at all.[76]

Jerry A. Fodor and Zenon W. Pylyshyn elucidate the need for logically articulating the systematicity of linguistic capabilities and general reasoning, suggesting that mental representations should presuppose a language of thought:

> What does it mean to say that thought is systematic? Well, just as you don't find people who can understand the sentence 'John loves the girl' but not the sentence ''the girl loves John,' so too you don't find people who can *think the thought* that John loves the girl but can't think the thought that the girl loves John. Indeed, in the case of verbal organisms the systematicity of thought follows from the systematicity of language if you assume – as most psychologists do – that understanding a sentence involves entertaining the thought that it expresses; on that assumption, nobody could understand both of the sentences about John and the girl unless he were able to think of the thoughts about John and the girl. But now, if the ability to think that John loves the girl is intrinsically connected to the ability to think that the girl loves John, that fact will somehow be explained. For a representationalist the explanation is obvious. Entertaining thoughts require requires being in representational states (that is, it requires tokening mental representations). And, just as the systematicity of language shows that there must be structural relations between the sentence 'John loves the girl' and the sentence 'the girl loves John,' so the systematicity of thought shows that there must be structural relations between the mental representation that corresponds to the thought that John loves the girl and the mental representation that corresponds to the thought that the girl loves John, namely, the two mental representations, like the two sentences must be made of the same part. But if this explanation is right, then mental representations have internal structure and there is a language of thought.[77]

Although Fodor and Pylyshyn's blanket assertion that the "architecture of mind is not a connectionist network," is certainly debatable, there is much merit to their argument that one cannot do without a "language of thought," even if the language of thought happens to be only a mental approximation to make sense of neurological activity.

None of the above-stated criticisms directed against eliminative materialism diminish the value of connectionism as a rather useful model of cognition. However, caution should be exercised before making sweeping generalizations about the nature of mental reality. The eliminative materialist project of Churchland and associates, are strikingly reminiscent of wholesale reductionisms made by behaviorists, especially B. F. Skinner. Here are a few samples of Skinner's rhetorical legacy that came to dominate eliminative and behavioristic/materialistic accounts of psychology:

> We change the relative strengths of responses by differential reinforcements of alternative courses of action; we do not change something called a preference. We change the probability of an act by changing a condition of deprivation or aversive stimulation; we do not change a need. *We reinforce behavior in particular ways; we do not give a person a purpose or an intention.*[78]

> Beliefs, preferences, perceptions, needs, purposes, and opinions are possessions of autonomous man which are said to change when we change minds.[79]

> Mentalism kept away from the external antecedent events which might have explained behavior, by seeming to supply an alternative explanation. Methodological behaviorism did just the reverse: by dealing exclusively with external antecedent events it turned away from self-observation and self-knowledge. Radical behaviorism restores some kind of balance. It does not insist upon truth by agreement and can therefore consider events taking place in the private world within the skin. It does not call these events unobservable, and it does not dismiss them as subjective. It simply questions the nature of the object observed and the reliability of the observations. The position can be stated as follows: what is felt or introspectively observed is not some nonphysical world of consciousness, mind or mental life but the observer's own body.[80]

Some forms of human behavior do indeed endorse Skinner's model of cognition based on the principle of reinforcement, as elucidated by the theory of operant conditioning. However, Skinner makes sweeping generalizations to explain away concepts such as freedom, dignity of humans, and other essentialist categories as mental fictions: "the autonomous man – the inner man, the homunculus, the possessing demon, the man defended by the literature of dignity and freedom" must be abolished.[81] To preempt criticism, Skinner adds "science does not dehumanize man; it dehomunculizes him…only by *dispossessing* him can we turn to the *real* causes of human behavior. Only then can we turn from the *inferred* to the observed, from the miraculous to the natural, from the *inaccessible* to the manipulable."[82]

Skinner's project to exorcise the ghosts of metaphysical categories such as beliefs, desires, intentions and so forth is quite ironically, self-contradictory. It is inconceivable how one can pronounce the destruction of metaphysics, without resorting to some conceptual 'metaphysical' approximations themselves. Metaphysical entities such as beliefs, desires among others are linguistic descriptions of some sort or the other, even if their referents are non-linguistic abstractions. One way to work around the radical reductionist program of Skinner is to replace the term 'metaphysical descriptions' with the term 'linguistic descriptions,' presuming that these descriptions stand for metaphysical entities. For instance, Skinner and behaviorists often use the term 'reinforcement.' On one hand, the term evokes a rather clinical and 'value-free' use of language, especially when one refers to reinforcement as a process whereby a particular behavior is strengthened by the adjustment of connection weights. One could even use 'value-free' numbers to represent the connections. Such a gesture would seem non-intentional and merely descriptive from on objectivist standpoint. But, on the other hand, the question arises whether the mere use of clinical language (terms such as stimuli, response, reinforcement) could in effect dissociate the intentional agent (who is manipulating the variables) from his or her inquiry? And by the same token, does the clinical language employed to describe the process of reinforcement in itself negate the possibility that the notion of reinforcement is also merely another type of a linguistic description, entailing some sort of a 'metaphysical' belief or the other? Having said so, I can anticipate the eliminative behaviorists' counter-arguing that linguistic descriptions themselves are effected through neurological processes in the brain. But, and but, the very explanation of neurological phenomena is impossible without using the resources of language, whether the nature of these descriptions are clinical or non-clinical. And even the very development of a clinical language presupposes a conscious selection of words that in turn entails the presence of active agents, not to mention Agent Skinner's own clinical desire to eliminate intentionality and agency from the behaviorist's picture of the world.

I do recognize that even my very modest counter-arguments are impossible without brain processes, and by the same token the enunciation of brain processes is impossible without language either. And language is impossible without brain processes, and the enunciation of brain processes is impossible without the use of language. Therefore, one could argue interminably to infinite regress with reasoning of the 'chicken or the egg' type. To make it brief, one cannot simply wish away the role of the agent or dissociate the knower from the known even from behaviorists accounts since the very enunciation of the idea of 'reinforcement' (despite its purported empirical support) is very much a linguistic category as well, a category subjected to the same linguistic processes that folk psychologists, humanists, or any other theory is subjected to.

The same criticism that applies to Skinner can apply to eliminative materialism as a whole, including Churchland's eagerness to demystify all common sense accounts of mind. Language is purposive, intentional, agential among many others – even if the account of folk psychology is radically false and misleading, neuroscientific accounts of cognition cannot escape the thrall of language, and will be therefore subjected to the same logical and sentential scrutiny that other theories normally face. Churchland is perhaps right in saying that:

> Thus are we led rather swiftly to the idea that there is a level of representation *beneath* the level of sentential or propositional attitudes, and to the correlative idea that there is a learning dynamic that operates primarily on sublinguistic factors. This idea is reinforced by reflection on the problem of cognition and learning in nonhuman animals, none of which appear to have the benefit of language, either the external speech or the internal structure, but all of which engage in sophisticated cognition. Perhaps their cognition proceeds entirely without benefit of any system for processing sentencelike representations.[83]

The point of my agreement with Churchland is with regard to the inevitability of neural material vis-à-vis cognition. However, what is ironic perhaps is Churchland's assumption that cognition is not based on 'sentencelike' propositions. If cognition is not based on sentencelike propositions, what about the representation of cognition *qua* cognition, a cognitive activity in itself? Can one resort to representing the representation of cognition without resorting to sentencelike structures? How can we know anything about the purported nonlinguistic nature of cognition without resorting to sentencelike structures?

Furthermore, the rapidity with which eliminative materialists are sweeping away intentions, beliefs, desires among other things that make us human in the guise of connectionism is rather troublesome. Such stark neurological determinism portrays a bleak fatalistic view of humanity where humans are deprived of their responsibility for ethical action. Furthermore, many societies are promoting ruthless capitalist technocracies whereby humans are becoming increasingly dispensable, even while a few profit mongers reap the accompanying economic rewards that come with such dehumanization. None of the criticisms directed against Skinner nor Churchland imply that they necessarily endorse such a perspective of a bleak and valueless world, but it is important to pay attention to question of values and ethics even while talking about serious scientific issues, since the façade of valuelessness, is in itself a value of some sort or the other.

Rhetorical Reflections

Artificial Intelligence claims to demystify the mind and places all mental processes within the purview of scientific investigation. The obvious advantage of an empirical approach lies in addressing some of the hitherto unanswerable questions about human cognition about the nature of human thought and learning, by taking cognition away from mentalism and reconstruing it in neurobiological physicalist terms. Connectionist AI stakes its claims on neurobiological grounds by asserting that a) the mind is a connection system and b) cognition is primarily a neurological process, and not a sentential or common sensical one and c) replacing the metaphysical 'mind' with the neurobiological 'brain.'

From a rhetorical perspective, matter to the AI community is an all-consuming terministic screen, and mind is either seen as an epiphenomenal or emergent property whose subjective effects could remain untouched, since the microstructure of the brain and the properties associated with cognition became more important than internal and subjective mentalistic concepts. Scientific disengagement gained precedence over subjective and emotivist accounts of cognition – for instance, instead of worrying about what one feels about the aesthetics of the rainbow, study the neuronal interaction in the brain that makes the perception of the rainbow possible. Human experience such as inner subjectivities of beliefs, desires, intentions and even logical reasoning are subordinated to studying the hardwiring of the brain. In Paul Churchland's words:

> Guided by our understanding of those internal structures, we manage to construct a new system of verbal communication entirely distinct from natural language, with a new and more powerful combinatorial grammar over novel elements over novel combinations with exotic properties...Once constructed, this "language" proves to be learnable; it has the power projected; and in two generations it has swept the planet. Everyone uses the new system. The syntactic forms and semantic categories of so-called "natural" language disappear entirely. And with them disappear the propositional attitudes of Folk Psychology, displaced by a more revealing scheme ... Folk Psychology suffers elimination.[84]

The brain's microstructure wherein a massive parallel processor performs computations through the inhibition and facilitation of signals, with connection weights constitutes the new language of cognition. Even without disputing the scientific claims of connectionism, rhetoricians could pose a question regarding the place or vantage point from which eliminative connectionists are advancing their claims. The view proposed is one of detached objectivity and neural realism untouched by natural language nor a priori logical reasoning nor participant observations.

An incisive commentary on the generative, philosophical fiction expressing such detachment is best expressed in Steve Woolgar's words:

> Prior to the modern age of the sociology of science, science was generally construed as a product untouched by the social and cultural forces often accepted as bearing upon all other (lesser) forms of knowledge. There was no sociology of scientific knowledge because the scientific production of representations was reckoned to involve no social factors. *The relationship between object and representation was treated as a black box by sociologists (and other theorists of knowledge). As long as scientific investigation could be explained in terms of cognition, it was neither necessary nor desirable to use a sociological explanans: R (representation) followed from O (object) as long as the appropriate thought processes (cognitive activities) were in operation. A second related perspective did admit the relevance of social factors, but only in circumstances where erroneous scientific knowledge was seen to result. The perspective is tantamount to a partial opening of the black box. The question posed – what led to the generation of this or that incorrect representation?- allowed for sociological speculation, but only in instances where something went wrong...The otherwise smooth cognitive operation of cognitive processes in the process of connecting objects and their representations was thought to have been distorted, deflected by the intervention of things social.*[85]

From the perspective of a sociology or a rhetoric of science, cognitive concepts are not seen as asocial, disengaged and disinterested reflections of the mental world as they are, but as necessary constructs mediated in a social world, to make sense of the mental world. Robert D. McPhee illustrates this position fairly well by stating that

> cognitive constructs are not only learned in organized social practices; they are often primarily social in their very nature. They are not primarily unseen, mysterious processes hidden in the cognitive system; they are things we do in the world, or features of those things. They can be found stated explicitly, practiced regularly out in the open, available to be learned, written up as a textbook or computer program, or "worked through" by a group."[86]

The idea that thought processes are sociological can seem scandalous or too radical to pure cognitivists, since opening the doors to sociological or rhetorical factors might appear to undermine notions of simon-pure objectivity.

The 'social' appears only as an after-thought in pure cognitivism, which is best illustrated in John O. Greene's overview of cognitivism:

> From a cognitive perspective, behavior is to be explained by reference to mental operations which produced it. No collective has ever performed pattern recognition on information residing in iconic short-term sensory storage or generated a single efferent signal. In other words, groups do not perform meaning analysis or exhibit muscle movements; people do. Similarly, in the view of cognitivism, the actions of others do not produce behavior, the meaning assigned to that behavior via cognitive processes does; social rules do not produce behavior, long-term memory representations of action-outcome contingencies do; social regularities and instructions do not produce behavior, learning processes operating on cognitively-detected patterns do. One advantage of cognitivism is that the boundaries of the cognitive system are clearly demarcated such that we can readily distinguish that which is cognitive from what is not. Once light and sound waves enter iconic and auditory storage, they cease being "social" and become "cognitive." Thus, social factors cannot override or suppress cognitive processes; they give way to cognitive processes.[87]

From the perspective of a cognitivist, social experience may be too broad a notion due to the level of difficulty in breaking up what constitutes the 'social' into discrete, quantifiable variables. On the other hand, recasting social experiences in terms of a clinical language that employs terms such as stimuli or input is a more feasible undertaking for the cognitivist project. In connectionist parlance, mental experience should be seen as a result of neural interconnections and wirings in the brain since the mind is a neural connection system after all. However, the purported delineation and subsequent delimitation of inquiry in terms of discrete, neurological variables eclipses other factors that come into play. I will at least identify two interrelated factors: a) the agential rooted in a sociology of science and b) the rhetorical rooted in the dialogic interplay between agent and language.

The agential factor is based on the premise that there are live, active agents who are active participants in constructing epistemological coalitions; where paradigms are either pitted against each other or engage in cooperative endeavors by clarifying, contributing and enhancing understanding of a particular subject. For instance, the rhetorical situation facilitating the connectionist paradigm, as discussed in the preceding sections, owes much to the confluence of intellectual developments and the dynamic interaction of agents who play an influential role in developing a particular school of thought or orientation. Needless to say, in spite of the role of human actors it is customary to render the presence of human subjects as extraneous to the inquiry. The use of non-emotive, clinical language in scientific inquiry highlights at least two things: a) the actual and procured knowledge from a particular experiment or hypothesis, presented in such a manner that is considered devoid of human intentionality and b) low-key verbal acknowledgment of individuals, institutions, spatial and temporal conditions that facilitate the furtherance of the inquiry. However, in the actual presentation of a technical report, the agent is imperceptibly obliterated by the use of a non-intentional, technical vocabulary. Given this backdrop, it is understandable how one could easily be led astray into thinking that the mere use of technical language implies a value-free orientation since a purported value-free orientation conveys the impression that there is a distinct lack of agential constructions. This is not to say that science eliminates the human agent, but that the agent is rendered as a mere passive spectator, timidly nodding to Parmenides and Plato that reality is an inflexible given.

Specific to connectionism, the agential role is manifest in discursive communities who ascribe to the common theme that 'computers are like brain,' and that neural-like implementations of mental activity are possible in artificial neural nets, starting progressively from simpler models to more complex models that simulate the more difficult aspects of human cognition. The neural inspiration of connectionist models can be gleaned in the paradigmatic shift from the mathematical/physics model of 'external' reality to more organic and biological models of 'external' reality. Walter McCulloch's remarks about translating even physics in terms of neurology is illuminating:

> Let us now compel our physicist to account for himself as part of the physical world. In all fairness, he must stick to his own rules and show in terms of mass, energy, space and time how it comes about that he creates theoretical physics. He must then become a neurophysiologist (that is what happens to me), but in so doing he will be compelled to answer whether theoretical physics is something which he can discuss in terms of neurophysiology (and that is what happened to me). To answer "No" is to remain a physicist undefiled. To answer "Yes" is to become a metaphysician – or so I am told.[88]

The preeminence accorded to biological is rhetorically compelling, in Henri Poincare's words, "the biologist has been led instinctively to regard the cell as more interesting than the whole animal, and the event has proved him right, since cells belonging to the most diverse organisms have greater resemblances. …than the organisms themselves."[89] Since AI entails the blurring of cognitive boundaries between artificial and natural systems, McCulloch and Poincare's comments are rhetorically compelling – McCulloch's because he foregrounds the importance of neurophysiology in perception, and Poincare because he articulates how biology turns to the cellular or microstructures of organisms drawing similarities between and across species. From that perspective, connectionist models of cognition attempt to blur the cognitive boundaries between humans and machines, by increasingly turning to the biology of the brain. Such a turn is not a mere fortuitous accident, but an intentional move with widespread growing cultural and intellectual interest in studying microstructures of biological systems, be it the genome project, neurons or the like. Invoking the biological argument, when placed in a wider context tends to rhetorically validate the emergence of connectionism, although evidence is still sketchy.

In short, active human agents are involved behind uncovering biological evidence and at the same time, constructing models as depictions of reality. Models are both scientific and rhetorical constructions as well. They are scientific in that at some level of description there is an attempt to garner empirical evidence, and they are rhetorical, in that the descriptions of reality are linguistic statements exercised by a conscious and scrupulous selection of words, arguments and so forth. Agreeing with Herbert Blumer, an agential account of scientific discourse would declare, "it is impossible to cite a single instance of a characterization of the world of reality that is not cast in the form of human imagery."[90] Human imagery is important even for clinical and seemingly 'non-intentional' descriptions. John Dewey summary of the role of language in the characterization of perceptions is perhaps apt:

> Without language, the qualities of organic action that are feelings are pains, pleasures, odors, colors, noises, tones, only potentially and proleptically. With language they are discriminated and identified. They are then 'objectified,' they are immediate traits of things...The qualities were never 'in' the

> organism; they always were qualities of interactions in which both extraorganic things and organisms partake. When named, they enable identification and discrimination of things to take place as means in a future course of inclusive interaction.[91]

From the discussion so far, one can at least acknowledge the central role that discourse plays in shaping our understanding of the physical and the mental world.

The discourse of connectionism accords centrality to the brain, or the formal approximation of its processes in artificial neural nets. In other words, what happens to be its unique strength is its willingness to take the route of neural realism, or at least have neural realism as its ultimate goal. Yet, its strength happens to be its weakness as well – in that cognition is bracketed off, as an entity separate from the external world and a world of social actors. However, from the perspective of modeling, it is easier to do this than to try to attempt to model societal interactions as a whole, which would be, prove to be cumbersome. Yet, the realm of the 'social' is deflected in most psycho-biological accounts. Needless to say, the tremendous promise that connectionist models are showing only goes to show that connectionism is here to stay.

The Biological Metaphor

David E. Rumelhart considers connectionist models of cognition to be neurally inspired, since computation is carried out through simple interactions among units that are considered equivalent of neurons:

> The basic strategy of the connectionist approach is to take as its fundamental process unit something close to an abstract neuron. We imagine that computation is carried out through simple interaction among such processing units. Essentially the idea is that these processing elements communicate by sending numbers along the lines that connect the processing elements. This identification already provides some interesting constraints on the kinds of algorithms that might underlie human intelligence. The operations in our

> models then can best be described as "neurally-inspired." *How does the replacement of the computer metaphor with the brain metaphor as model of mind affect our thinking?* This change in orientation leads us to a number of considerations that further inform and constrain our model-building efforts.[92]

Rumelhart acknowledges the conceptual salience of the 'brain' metaphor in generating new cognitive concepts, even while acknowledging that technology has provided a series of stock metaphors with which to view physical and mental realities. It is worth quoting Rumelhart at length:

> I recall vividly a class I taught some fifteen years ago in which I outlined the then current-view of the cognitive system. A particularly skeptical student challenged my account, with its reliance on concepts drawn computer science and artificial intelligence, with the question of whether I thought my theories would be different if it had happened that our computers were parallel instead of serial. My response as I recall, was to concede that our theories might very well be different...I pointed out that the inspiration of our theories and our understanding of abstract phenomena always is based on our experience with the technology of the time. I pointed out that Aristotle had a wax tablet theory of memory, that Leibniz saw the universe as clockworks, that Freud used a hydraulic model of libido flowing through the system, and that the telephone-switchboard model of intelligence had played an important role as well. The theories posited by those of the previous generations had, I suggested, been useful in spite of the fact that they were based on the metaphors of their time. Therefore, I argued, it was natural that in our generation – the generation of the serial computer – we should draw our insights from analogies with the most advanced technological developments of our time.[93]

The computer metaphor generated valuable cognitive concepts because "we can use the computer to simulate systems with which we *wish* to have experience and thereby provide a source of experience that can be drawn upon in giving us new metaphors and new insights into how mental operations might be accomplished."[94] Rumelhart further explains, "it is this use of the computer that the connectionists have employed. The architecture that we are exploring is not one based on von Neumann architecture of our current generation of computers but rather an architecture based on considerations of how brains themselves might function."[95] Therefore, "our goal in short is to replace the computer metaphor with the brain metaphor."[96]

A brief comparison of natural and artificial neurons is helpful, in terms of explaining how the mapping occurs:

Nervous System	Artificial neural net
Neuron	Abstract neuron or processing element
Dendrites	Incoming connections
Cell body	Activation level, transfer, activation and output function
Spike	Output of a node
Axon	Connection to other neurons
Synapses	Connection strengths and weights
Spike Propagation	Propagation rule

For instance, a connectionist commenting on a system engaging in pattern-recognition would probably effect the following mapping:

Human eye	Machine Implementations
Visual Sensors	Camera
Proximity/Distance	IR, ultrasound, laser range-finder sensor and so forth

The brain and all its related activities constitutes the source domain and the artificial neural network serves as the target domain whereby mapping takes place from one domain onto another. The isomorphism between the biological neuron and the artificial neuron is a metaphoric equivalence and the dendrites in the natural neuron purportedly correspond to the cell connections in artificial nets, where

Synapses (are isomorphic to) Weights and

Axons (are isomorphic to) Outputs.

In effect, if one were to build an artificial neural network it is important to specify at least the following, all of which have its neurological counterparts -

Specifying the characteristics of the artificial neuron (as nodes, processing elements, neurons or model neurons synonymously) in a neuroscientific vocabulary which explains the relationship of the nodes with the incoming inputs, how the node sums the input, how the inputs are transformed into a level of activation, how the inputs are transformed into outputs that are transmitted along the axon.

Specifying what nodes are connected and the direction in which they are connected.

Specifying how a given activation traveling along an axon transmits to the neurons to which the axons are connected (propagation rule).

Specifying the connection strengths between neurons and how they are subjected to change over time (learning rule).

All of these assumptions are based on establishing a correlation between the neurophysiological properties of the brain and the cognitive properties of the machine implements. The differences in 'hardware' could also lead to differences in processes, because an insistence that dissimilitude of 'hardware' is irrelevant to thinking only presupposes some sort of Platonic conception of 'thinking' - where distinctive processes are ignored, for the sake of achieving certain goals. And if processes are ignored, then the conception of an entity being a model of something else simply breaks down. Therefore, the idea of neural modeling hinges on a theory of representation – namely, to see the mind as a vector space. Neural networks may be seen as a schematization of the organic nervous system where both the biological system and the artificial system may have similarities with reference to a structure of interconnected entities and also mechanisms for adaptations (learning algorithms), but the functional similarities by themselves do not necessarily mean that the analog established between the two is literal – not to mention the differences in architecture and biological material.

However, artificial neural networks are seen as effective approximations of the central nervous system to the extent that one could reconceptualize how psychology has to be done:

> Connectionist models are leading to a reconceptualization of key psychological issues, such as the nature of the representation of knowledge...One traditional approach to such issues treat knowledge as a body of rules that are consulted by processing mechanisms in the course of processing; in connectionist models, such knowledge is represented, often in widely distributed form, in the connections among the processing units.[97]

By reconceiving the mind primarily in terms of vector spaces, one of the hopes is to dispel folk psychology's 'language of thought' or internal brain language and also do away with the idea of 'mentalese' language like causation – example being, "I carry an umbrella because I *believe* it might rain." An incoming vector activates the relevant portions of the network by "virtue of its own vectorial makeup."[98] In Churchland's words:

> The old problem of how to retrieve relevant information is transformed by the realization that it does not need to be retrieved. Information is stored in brain-like networks in the global pattern of their synaptic weights. An incoming

> vector activates the relevant portions, dimensions, and subspaces of the trained network by virtue of its own vectorial makeup. Even an incomplete version of a given vector (that is, one with several elements missing) will often provoke essentially the same response as the complete vector by reason of its relevant similarity.[99]

Thinking is primarily information-processing in terms of inputs, hidden units and outputs in a vector space. Thinking of 'thinking' as only vectorial activation in a neurological substratum of the brain necessitates a different epistemology of mind. However, in order to do so, one has to import the vocabulary of information technologies and redescribe the brain primarily in terms of the vocabulary of transmission technologies, and then map the neuroscientific vocabulary onto the artificial net. If information is embodied *only* in the discrete set of processed signals, things such as beliefs, desires, values should be reconstrued in the neurological language of activation vectors. If the 'mind' is *just* a vector space, ideas such as mental causation, language of thought and a whole barrage of mentalistic vocabulary are ejected, and by the same token, the neural models should be seen not merely as the implementation of cognitive models, but as actual reflections of human cognition. Any difficulties that might complicate the picture, should be primarily treated as a complexity issue – namely, more complex networks can eventually resolve the issue of 'ambiguous' outputs.

A little bit of reflection on the nature of the verisimilitude between artificial nets and the nervous system it claims to replicate is instructive. Frank Rosenblatt's ruminations on the nature of perceptrons are perhaps pertinent even today with regards to the looming question regarding the accomplishment of neural realism:

> Perceptrons are not intended to serve as detailed copies of any actual nervous system. They're simplified networks, designed to permit the study of lawful relationships between the organization of a nerve net, the organization of its environment, and the "psychological" performances of which it is capable. Perceptrons might actually correspond to parts of more extended networks and biological systems; in this case, the results obtained will be directly applicable. More likely they represent extreme simplifications of the central nervous system, in which some properties are exaggerated and others suppressed....The main strength of this approach is that it permits meaningful questions to be asked and answered about particular types of organizations, hypothetical memory mechanisms, and neural models.[100]

David E. Rumelhart, James L. McClelland and the PDP group similarly suggest:

> To be sure, to the extent that our models are directly relevant to brains, they are at best coarse approximations of the details of neurophysiological processing. Indeed, many of our models are clearly intended to fall at a level between the macrostructure of cognition and the details of neurophysiology. Now, we do understand that some of our approximations may have ramifications for the cognitive phenomena which form our major area of interest, we may be missing out on certain aspects of brain function that would make the difference between an accurate account of cognitive-level phenomenon and a poor approximation...A neuroscientist might be concerned about the ambiguity inherent in the fact that many of the mechanisms we have postulated could be implemented in different ways. From our point of view, though, this is not a serious problem...But since our primary concern is with the computation themselves, rather than detailed neural implementation of these computations, we are willing to be instructed by neuroscientists...Nevertheless, we have chosen a level of approximation which seems to be most fruitful, given our goal of understanding the human information processing system.[101]

The 'brain metaphor' is intended to be a model to shed light on the human information processing system. The brain metaphor is a biological metaphor in that it seeks "understanding of intelligence directly in terms of biological function, rather than indirectly through molar levels of processing."[102] The biological function in connectionist models are largely construed in terms of neurological function that means that other functions are ignored. Furthermore, sociological, environmental and other factors are hardly addressed in connectionist models. The communicative pattern of neurons is largely based on input-output models. Limitations of input-output models of information/communication have been extensively critiqued in a vast body of literature.

Notes

[1] Claude Bernard, *An Introduction to the study of Experimental Evidence*, trans. Henry Copley Green (New York: Dover Publications, 1957), 78.
[2] cited in John F. Fulton, *Selected Readings in the History of Physiology* (Springfield, Illinois: Charles C. Thomas, 1930), 308.
[923] Bernard, *An Introduction to the study of Experimental Evidence*, 78-79.
[4] Bernard, *An Introduction to the study of Experimental Evidence*, 78.
[5] Bernard, *An Introduction to the study of Experimental Evidence*, 90.
[6] Edward Lee Thorndike, *Human Learning* (New York: Appleton-Century Crofts, 1931), 122.
[7] Arthur C. Guyton, *Basic Neuroscience* (Philadelphia: W. B. Saunders company, 1991), 86.
[8] Guyton, *Basic Neuroscience*, 87.
[9] Edward L. Thorndike, *Selected Writings From a Connectionist's Psychology* (New York: Appleton-Century-Crofts, 1949), 17.
[10] Edward L. Thorndike, "Why Study Animal Psychology?" in *Comparative Psychology*, ed. F. A. Moss (New York: Prentice-Hall, 1942), 1-6 (5).
[11] Claude E. Shannon, *A Symbolic Analysis of relay and switching circuits*, Master's Thesis, Massachusetts Institute of Technology; published in *Transactions of the American Institute* 57 (1938): 1-11.
[12] Claude E. Shannon and Warren Weaver, *The Mathematical Theory of Communication* (Urbana: University of Illinois Press, 1962).
[13] Walter S. McCulloch and Walter H. Pitts, "A logical calculus of the ideas immanent in nervous activity," in *The Philosophy of Artificial Intelligence*, ed. Maragaret A. Boden (Oxford: Oxford University Press, 1990), 22-39.
[14] McCulloch and Pitts, "A logical calculus of the ideas immanent in nervous activity," 37.
[15] McCulloch and Pitts, "A logical calculus of the ideas immanent in nervous activity," 25.
[16] R. Morenzo-Diaz and J. Mira, "Logic and Neural Nets: Variations on Themes by W. S. McCulloch," in *Brain Processes, Theories and Models: An International Conference in the Honor of W. S. McCulloch 25 years after His Death*, ed. Roberto Moreno-Diaz and José Mira-Mira (Cambridge, Massachusetts: MIT Press, 1996), 24-36 (25).
[17] Morenzo-Diaz and Mira, "Logic and Neural Nets," 24.
[18] J. Mira, "Reverse Neurophysiology: The "Embodiments of Mind" Revisited," in *Brain Processes. Theories and Models: An International Conference in the Honor of W.S.McCulloch 25 years after His Death* , ed. Roberto Moreno-Diaz and José Mira-Mira (Cambridge, Massachusetts: MIT Press, 1996), 37-49 (38).
[19] Mira, "Reverse Neurophysiology," 37-38. * Phylogeny and ontogeny refer to the developmental history of groups of organism and an individual organism respectively.
[20] cited in Michael K. Weir, "Putting the Mind Inside the Head," in *Brain Processes, Theories and Models: An International Conference in the Honor of W.S.McCulloch 25 years after His Death* , ed. Roberto Moreno-Diaz and José Mira-Mira (Cambridge, Massachusetts: MIT Press, 1996), 54-63 (54).
[21] Mira, "Reverse Neurophysiology," 43.
[22] John Von Neumann, "First Draft of a Report on the EDVAC," in *The Origins of Digital Computers*, ed. Brian Randell (New York: Springer-Verlag, 1973), 355-364 (362).

•

[23] Walter H. Pitts and Warren S. McCulloch, "How we know universals, the perception of auditory and visual forms," *Bulletin of Mathematics and Biophysics* 9 (1947): 127-147.
[24] J. Y. Lettvin, H. Maturana, W. S. McCulloch and W. H. Pitts, "What the frog's eye tells the frog's brain," *Proceedings of the IRE* 47 (1959): 1940-1951.
[25] Lettvin et al, "What the frog's eye tells the frog's brain," 1949.
[26] Mira, "Reverse Neurophysiology."
[27] Norbert Wiener, *Cybernetics, or Control and Communication in the Animal and the Machine* (New York: Wiley, 1948).
[28] Wiener, *Cybernetics, or Control and Communication in the Animal and the Machine*, 14.
[29] K. M. Sayre, "Cybernetics," in *Philosophy of Science, Logic and Mathematics in the 20th century*, ed. Stuart G. Shanker (Routledge: New York , 1996), 292-314 (312).
[30] cited in Sayre, "Cybernetics," 312.
[31] Sayre, "Cybernetics," 311.
[32] McCulloch and Pitts, "A logical calculus of the ideas immanent in nervous activity."
[33] McCulloch and Pitts, "A logical calculus of the ideas immanent in nervous activity," 24.
[34] F. Rosenblatt, "The Perceptron: A Probabilistic model for information storage and organization in the brain," *Psychological Review* 65, no. 6 (1958): 386-408 (387-388).
[35] Rosenblatt, "The Perceptron," 386.
[36] Rosenblatt, "The Perceptron," 388-389.
[37] Rosenblatt, "The Perceptron," 406.
[38] Marvin Minsky and Seymour Papert, *Perceptrons: An Introduction to Computational Geometry* (Cambridge, Massachusetts: MIT Press, 1969), 19.
[39] F. Rosenblatt, *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms* (Washington, D.C: Spartan Books, 1962), 386.
[40] cited in Margaret Boden, *The Philosophy of Artificial Intelligence* (Oxford: Oxford University Press, 1990), 327.
[41] Marvin Minsky, "Connectionist Models and their Prospects," in *Connectionist Models and their implications: Readings from Cognitive Science*, eds. David Waltz and Jerome A. Feldman (Norwood, New Jersey: Ablex Publishing Corporation, 1988), vii-xvi (viii).
[42] Paul Smolensky, "Connectionist Modeling: Neural Computation/Mental Connections," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: MIT Press, 1997), 233-250 (236).
[43] Smolensky, "Connectionist Modeling," 233.
[44] Guyton, *Basic Neuroscience*, 108.
[45] Guyton, *Basic Neuroscience*, 108.
[46] Guyton, *Basic Neuroscience*, 109.
[48] Paul M. Churchland, "On the Nature of Theories: A Neurocomputational Perspective," in in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: MIT Press, 1997).
[49] Churchland, "On the Nature of Theories," 258.
[50] John Haugeland, "What is Mind Design?" in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: MIT Press, 1997), 1-28 (21-22).
[51] Smolensky, "Connectionist Modeling," 237.
[52] Smolensky, "Connectionist Modeling," 237.
[53] Smolensky, "Connectionist Modeling," 238.
[54] cited in Jerry A. Fodor and Zenon W. Pylyshyn, "Connectionism and Cognitive Architecture: A Critical Analysis," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: MIT Press, 1997), 309-350 (341).
[55] David E. Rumelhart, James L. McClelland and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. I: Foundations* (Cambridge, Massachusetts: MIT Press, 1986), 10.
[56] Rumelhart, McClelland et al, *Parallel Distributed Processing Vol. I*, 11.

[57] Rumelhart, McClelland et al, *Parallel Distributed Processing Vol. I,* 12.
[58] Rumelhart, McClelland et al, *Parallel Distributed Processing Vol. I,* 32.
[59] Rumelhart, McClelland et al, *Parallel Distributed Processing Vol. I,* 32.
[60] David E. Rumelhart, "The Architecture of Mind: A Connectionist Approach," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence,* ed. John Haugeland (Cambridge, Massachusetts: MIT Press, 1997), 205-232 (209).
[61] Rumelhart, "The Architecture of Mind," 209.
[62] T. Sejnowski and C. Rosenberg, "Parallel networks that Learn to Pronounce English Text," *Complex Systems* 1 (1987): 145-168.
[63] James L. McClelland, David E. Rumelhart and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. II: Psychological and Biological Models* (Cambridge, Massachusetts: MIT Press, 1986).
[64] J. L. Elman, "Distributed Representations, Simple Recurrent Networks, and Grammatical Structure," in *Connectionist Approaches to Language Learning* (Dordrecht, Netherlands: Kluwer, 1991), 91-122.
[65] D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning Internal Representations by Error Propagation," in , *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. I: Foundations,* eds. David E. Rumelhart, James L. McClelland and the PDP Research Group (Cambridge, Massachusetts: MIT Press, 1986), 318-362.
[66] Haugeland, "What is Mind Design?," 22.
[67] William Ramsey, Stephen Stich and Joseph Garon, "Connectionism, Eliminativism and the Future of Folk Psychology," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: MIT Press, 1997), 351-376; Paul M. Churchland, "Eliminative Materialism and the Propositional Attitudes," in *Problems in Mind: Readings in Contemporary Philosophy of Mind*, ed. Jack S. Crumley II (Mountain View, California: Mayfield Publishing Company, 2000), 184-197.
[67] Churchland, "Eliminative Materialism and the Propositional Attitudes," 184.
[68] Churhcland, "Eliminative Materialism and the Propositional Attitudes," 185.
[69] Churchland, "Eliminative Materialism and the Propositional Attitudes," 194.
[70] Ramsey, Stich and Garon, "Connectionism, Eliminativism and the Future of Folk Psychology," 356.
[71] Ramsey, Stich and Garon, "Connectionism, Eliminativism and the Future of Folk Psychology," 356.
[72] cited in Ramsey, Stich and Garon, "Connectionism, Eliminativism and the Future of Folk Psychology," 361.
[73] Churchland, "On the Nature of Theories," 283.
[74] Churchland, "On the Nature of Theories," 252.
[75] Jay F. Rosenberg, "Connectionism and Cognition," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: MIT Press, 1997), 293-308.
[76] Rosenberg, "Connectionism and Cognition," 307.
[77] Jerry A. Fodor and Zenon W. Pylyshyn, "Connectionism and Cognitive Architecture" A Critical Analysis," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: MIT Press, 1997), 309-350 (331-332).
[78] B. F. Skinner, *Beyond Freedom and Dignity* (New York: Knopf, 1971), 94.
[79] Skinner, *Beyond Freedom and Dignity*, 93.
[80] B. F. Skinner, "Excerpts from About Behaviorism," in *Problems in Mind: Readings in Contemporary Philosophy of Mind*, ed. Jack S. Crumley II (Mountain View, California: Mayfield Publishing Company, 2000), 59-67 (61).
[81] Skinner, *Beyond Freedom and Dignity*, 200.
[82] Skinner, *Beyond Freedom and Dignity*, 213-214.
[83] Churchland, "On the Nature of Theories," 253.
[84] Churchland, "Eliminative Materialism and the Propositional Attitudes," 195.
[85] Steve Woolgar, "Representation, Cognition and Self: What Hope for an Integration of Psychology and Sociology?," in *Ecologies of Knowledge: Work and Politics in Science and Technology*, ed. Susan Leigh Star (New York: State University of New York Press, 1995), 155-179 (159-160).

[86] Robert D. McPhee, "Cognitive Perspectives on Communication: Interpretive and Critical Responses," in *The Cognitive bases of Interprersonal Communication*, ed. D.E.Hewes (Hillsdale, NJ: Lawrence Erlbaum, 1995), 225-246 (236).
[87] John O. Greene, "Evaluating Cognitive Explanations of Communicative Phenomena," *Quarterly Journal of Speech* 70 (1984): 241-254 (243).
[88] cited in Heinz Von Foerster, "Metaphysics of an Experimental Epistemologist," in *Brain Processes, Theories and Models: An International Conference in the Honor of W. S. McCulloch 25 years after His Death*, ed. Roberto Moreno-Diaz and José Mira-Mira (Cambridge, Massachusetts: MIT Press, 1996), 3-10 (3).

[89] Henri Poincare, *Science and Method*, trans. Francis Maitland (New York: Dover Publications, 1952), 19.
[90] Herbert Blumer, *Symbolic Interactionism: Perspective and Method* (Englewood Cliffs, NJ: Prentice Hall, 1969), 22.
[91] Joseph Ratner, *Intelligence in the Modern World: John Dewey's Philosophy* (New York: Random House, 1939), 804.
[92] David E. Rumelhart, "The Architecture of Mind: A Connectionist Approach," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland (Cambridge, Massachusetts: MIT Press, 1997), 205-232 (207).
[93] Rumelhart, "The Architecture of Mind," 204.
[94] Rumelhart, "The Architecture of Mind," 206.
[95] Rumelhart, "The Architecture of Mind," 206.
[96] Rumelhart, "The Architecture of Mind," 206.
[97] Jay L. McClelland, Jerome A. Feldman, B. Adelson, Gordon H. Bower and Drew McDermot, *Connectionist Models and Cognitive Science: Goals, Directions and Implications.* Report to the National Science Foundation, June 1986 (12).
[98] Churchland, "On the Nature of Theories," 291.
[99] Churchland, "On the Nature of Theories," 291.
[100] Frank Rosenblatt, *Principles of Neurodynamics* (New York: Spartan, 1962), 28.
[101] David E. Rumelhart, James McClelland and the PDP research group, *Parallel Distributed Processing Vol 1: Foundations* (Cambridge, Massachusetts: MIT Press, 1986), 138.
[102] Sternberg, *Metaphors of Mind*, 162.

CHAPTER VII

CONCLUSION

This dissertation was a case study in the rhetoric of science. The approach here represents both continuity and a call for new directions in how rhetoric of science studiescould be done. It represents continuity in paying closer attention to how scientific epistemologies go hand-in-hand with discursive strategies. It provides new direction through the subject matter of this discussion by focusing on an emerging interdisciplinary area that cross-stitches neurosciences, psychology, computer science and philosophy. Studies in the rhetoric of science have so far laid the groundwork on articulating a case for the juxtaposition of rhetoric and science, a juxtaposition that seems counter-intuitive at first glance. Also, much attention has been paid to how scientific discourse is persuasive. However, this study calls us to do even more – namely, to look at rhetoric not merely as an entity that enters the discourse for the sake of winning assent, as scientists attempt to persuade their peers. The earlier approach concerns chiefly matters of style and presentation. However, if we look at rhetoric as an entity that is embedded in the discourse qua discourse, the scope of rhetoric can be extended to the actual generation of content as well. Although, there is much value in looking at how scientists persuade, there is also value in looking at how the discourse in itself is rhetorically generated.

What has been done well in the rhetoric of science so far is to show how scientific rhetors validate their messages before scientific audiences. What has not yet been done well is to show how the discourse in itself is generated rhetorically through the 'metaphor dependency' of the sciences. What this study hopes to do is to show one example of how rhetoric operates within the sciences as evidenced from the use of generative metaphors is directly involved with the development of models of mind. At the same time, it must be mentioned that the study (at the time of this writing) only reinforced the view that apt metaphors are theory generative. For future studies, rhetoricians conversant with other areas of the sciences such as quantum physics, genetics among others may also investigate whether metaphors have influenced the formation of particular epistemologies. If this is not always the case, one may very well abandon the hypothesis that all the sciences are metaphor-dependent in the spirit of intellectual honesty.

Another contribution or direction encouraged by this study is to look at the development of the metaphor through the rhetorical situation, therefore a reading of primary texts in technical literature is important, both from historical and rhetorical perspectives. However, it must be mentioned at the outset that this study invites analyses with more depth, in that it does not amplify on all the details of the discourse generated by the metaphors, in its quest for the bigger picture. Therefore, rhetorical studies in the future on AI can focus on specific topics such as – the frame problem (how do we represent), emotions, intentionality, agency, consciousness (how is the question of consciousness treated), robotics and so forth. In other words, this study can be seen as a tentative (and by no means exhaustive) exploration that is calling for more explorations on specific details. All these above-mentioned areas need more specific exploration from a rhetorical perspective.

Last but not the least, this study by using Burke's conception of a terministic screen hopes to situate Burke as a rhetorician of science as well. What does Burke teach us about science vis-à-vis the terministic screen? Burke tells us that our ways of knowing are also ways of not knowing. In fact, Burke's own position is quite consistent with the spirit behind Heisenberg's uncertainty principle, namely that what we focus on also becomes a point of deflection as well. If this extrapolation can be made, Burke can be seen as a philosopher or rhetorician of science who accepted the view of incompleteness as intrinsically related to the scientific enterprise.

Rear-view Synthesis

The field of AI is labyrinthine with multitudinous pathways. Therefore, the treatment accorded in this study is by no means exhaustive. The study set out to highlight the rhetorical situation and the role of metaphors as terministic screens in generating the discourse. Rhetoric plays a vibrant role in the form of persuasive thought-experiments and generative metaphors. Chapter I set the tone for the discussion by engaging in a preliminary exploration of the central thesis of this discussion, namely, that the discourse of mind implicit in AI is also rhetorically constructed as seen through: a) the emergent rhetorical situation of physicalism and b) the role of generative metaphors in creating concepts and theoretical models of cognition.

As stated earlier, AI as a cognitive science rests on at least three fundamental premises: a) there is a material basis to intelligence, b) intelligence can be simulated in non-human entities and c) the mind hitherto regarded as a prototype, in itself, is a biological machine. Set against this backdrop, chapter II articulates how AI fits in as a rhetoric of science and also argues how analogical reasoning in many ways forms the basis for tying together conceptions of human and machine intelligence through the use of metaphors that operate as terministic screens. The chapter concludes that metaphors form the basis by which an isomorphic mapping between the human mind and machines can be made. In the case of symbolic AI, the source domain is the machine by means of which a computational psychology of mind is effected. In the case of connectionist AI, the source domain is the human brain by means of which a neuroscientific psychology of machine intelligence is effected although exact neural realism is not necessarily the desired goal.

If metaphors are the generative tools that help in theory building, anthropomorphizing (Chapter III) pertains to the motivation behind the rhetoric of the discourse. Anthropomorphizing takes on two separate dimensions with reference to symbolic and connectionist AI. With regard to symbolic AI, anthropomorphizing can be seen as a desire to see the world through a rationalist paradigm, of which mathematical logic is the ultimate expression. Anthropomorphizing manifests itself in the desire to understand mental activity in a language where thought is treated as computation. Behind the idea of thought as computation lays the possibility of predictive power and causal generality accompanied with a precision that is found wanting in most conventional treatments of mind. Anthropomorphizing in connectionist AI can be seen as a desire to see the world through an empirical and more specifically biological (or biologized) paradigm. The shift from mathematics to biology is rhetorically salient since the latter speaks in the language of probability and not so much in the language of dogmatic certainty, by taking into fact that organic systems are more complex.

With connectionism, machines are treated as organic entities by interpreting artificial neural nets in terms of a neuroscientific vocabulary. This chapter argues that anthropomorphizing is based on at least two things: a) it underscores the intentionality of human agents in constructing theories, even in cases where these paradigms appear to be normative and ontological descriptions of an external reality independent of the observer (cases such as the origins of the universe may fall under this realm) and b) it operates even in conditions where it seems that theories appear to have a life of their own (in explaining the natural or social world) are themselves linguistic and rhetorical impositions of the human subject trying to understand the world in terms of orientations most amenable. Chapters II and III may be seen primarily as theoretical chapters engaging trying to one hand, bring out the rhetoric in the discourse of mind (presupposed by AI) and on the other hand, commenting on the motivation behind the individual rhetorics.

IV, V and VI may be seen as analysis chapters. Chapter IV deals with Alan Turing's idea of thought, which essentially discusses the mind as a discrete state machine with different states each of which is dependent on the preceding output and the incoming input signals. This chapter concludes that Turing equates thought with computation and that the imitation game should be seen as an outworking of the equivalence between thought and computation. In other words, if thought is computation the conception of intelligence can be easily extended to machines as well, since ultimately both humans and machines perform computation. Turing believed that computers would eventually pass the imitation game. On further reflection, one would not find Turing's confidence surprising because if one starts with the premise that thought is computation (and given the fact that machines do indeed compute) the range of activities that are at present difficult for machines (like carrying on a normal conversation) should primarily be described in computational terms, thus enhancing the feasibility for machines to accomplish these tasks. By the same token, what is considered for the present time non-computable should not be held against machines because humans are confronted by systemic limitations as well.

Chapter V treats the conception of thought in symbolic AI that is perhaps directly linked to Turing's conception of the mind. This chapter concludes that symbolic AI is instrumental in influencing a computational psychology of mind. The mind is treated as the target domain and the computer, the source domain – the mapping occurs in such a fashion that the mind is considered isomorphic with the machine. This chapter concludes that the metaphoric premises on which symbolic AI is based, is instrumental in construing thought as symbol manipulation and the mind as a mathematical, problem space where all mental activities can be defined as discrete, algorithmic procedures.

Chapter VI goes in the opposite direction, meaning that one starts with the brain as the source domain and look at machine as the target domain. A metaphoric mapping of neuroscientific vocabulary onto artificial nets occurs. This chapter concludes that 'thinking' according to connectionism rests on the idea of information processing, primarily in terms of inputs, hidden units and output units in a vector space. Thinking of 'thinking' as only vectorial activation is a neurological substratum necessitates a different epistemology of mind and that is exactly what connectionist psychology hopes to accomplish.

Terministic Screens and Metaphors of Mind

Since this discussion borrowed Burke's idea of a terministic screen, it is imperative to engage in some 'rounding out' as a tribute to perhaps one of the most influential rhetorical theorists of the twentieth century and also as a way of winding up this discussion. First, let us revisit the diagram proposed in chapter II to explain the relationship if any between symbolic and connectionist AI respectively.

Machine		Human Cognition
A	→	B
A_1 (Digital computer)	→	B_1 (Human cognition according to Symbolic AI)
A_2 (Artificial Net)	←	B_2 (Human cognition according to Connectionist AI)

B_2 replaces B_1

$B_2 \rightarrow A_1$

Essentially, the author seems to be suggesting that the shift from the mechanistic metaphor to the biological metaphor does not result in an abrupt shift from one form to another, but instead takes us full circle in that the biological metaphor is considered powerful enough to encompass symbolic AI as well. It must be mentioned though that the 'brain' does replace the 'mind' of folk psychology or symbolic AI. However, symbol processes can be seen as neuronal implementation – or at least that is what we are told, if one is to create a hybrid machine that can encompasses the strengths of symbolic and connectionist AI as well.

On one hand, the respective terministic screens played out differently in that they went in opposite directions – yet, on the other hand, as Kenneth Burke explains "language is intrinsically hortatory (a medium by which men can obtain the cooperation of one another)."[1] Both schools by its selection of facets of human cognition that are simulatable have also omitted details that seem to complicate the picture. However, what both schools share in common pertain to some type of Platonic idealization of mechanism and it is the common commitment to the idea of a machine that is capable of simulating human cognition.

Based on my reading of Burke's 'On Symbols and Society,' if Burke was still alive and was invited to comment on the creative tension between symbolic and connectionist AI, he would probably say, "such conflicts are clearly dialectical."[2] At the heart of the dialectical debate, is a revivified paradox of substance, and the idea of substance hinges on the idea of placement:

> To tell what a thing is, you place it in terms of something else. This idea of locating, or placing, is implicit in our very word for definition itself: to define, or determine a thing, is to mark its boundaries, hence to use terms that possess, implicitly at least contextual reference.[3]

However, as Burke continues:

> The word "substance," used to designate what a thing is, derives from a word designating something that a thing is not (placement). That is, though used to designate something within the thing, intrinsic to it, the word etymologically refers to something outside the thing, extrinsic to it. Or otherwise put: the word in its etymological origins would refer to an attribute of the thing's context, since that which supports or underlies a thing would be a part of the thing's

> context. And a thing's context, being outside or beyond the thing, would be something that the thing is not.[4]

A little bit of elaboration is necessary. The term 'machine' etymologically comes from the Latin *machina* and the Greek *mechos* originally refers to some sort of apparatus, conveyance, vehicle or some engine whose purpose was to serve as a means to an end. Traditionally, a machine has always been a means to an end, to fulfill some human need by efficiently accomplishing the task at hand. With the advent of AI, the term 'machine' is no longer a means to an end, but an end unto itself – in effect, its power has become so pervasive that it is purportedly a model for human minds as well. Therefore, more specifically by applying Burke one could say that both symbolic and connectionist AI attempts to grapple with cognition or the human mind by studying what the mind is not, namely the machine. Yet, the paradox of substance emerges when the machine becomes a mind or is treated as the mind.

Symbolic AI is quite content in starting with the 'machine' as the mind. Connectionism on the other hand, prefers to start with the human mind by recognizing the complexity involved with the minds and prefers building machines by studying the mind. However, in order to do so, the mind has to be some kind of a machine if it is serve as a prototype for intelligent machines. Therefore, what is common between the 'mechanistic' metaphor and the 'biological' metaphor is the idealization of the machine as the model for cognition.

Therefore the synthesis between the two, to Burke, might be 'dialectical.' So, we are back to the formulation that humans are after all machines. So, what does thinking of oneself as a machine entail? What are the advantages and the disadvantages? These are issues that one should let the kind reader to reflect upon.

Notes

[1] Kenneth Burke, *On Symbols and Society*, ed. Joseph R. Gusfield (Chicago: The University of Chicago Press, 1989), 74.
[2] Burke, *On Symbols and Society*, 202.
[3] Burke, *On Symbols and Society*, 237.
[4] Burke, *On Symbols and Society*, 237.

REFERENCES

Agre, Philip E. *Computation and Human Experience*. Cambridge: Cambridge University Press, 1997.

Anand, Vishwanathan. "More Questions than Answers," http://www.research.ibm.com/deepblue/home/may11/story_3.html, 1997.

Aristotle, *On Rhetoric: A Theory of Civic Discourse*, trans. George A. Kennedy. New York: Oxford University Press, 1991.

Auerbach, Erich. *Mimesis: The Representation of Reality in Western Literature*, trans. Willard R. Trask. Princeton, NJ: Princeton University Press, 1953.

Augustine. *The Confessions of Saint Augustine Book XI*, trans. Rex Warner, 274-276. New York: Signet, 2001.

Bernard, Claude. *An Introduction to the Study of Experimental Evidence*, trans. Henry Copley Green. New York: Dover Publications, 1957.

Bertalanffy, Ludwig von. *Perspectives on General Systems Theory*. New York: George Braziller, 1975.

Bethe, H. A. "Energy Production in the Stars," *Physical Review* 55(1939): 434-456.

Birkhoff, Garrett. "Mathematics and Psychology," *SIAM Review* 11(October, 1969): 429-469

Bitzer, Lloyd F. "The Rhetorical Situation." In *Contemporary Rhetorical Theory: A Reader*, ed. John Luis Lucaites, Celeste Michelle Condit and Sally Caudill, 217-225. New York: Guilford Press, 1999.

Black, Max. "More about Metaphor." In *Metaphor and Thought*, ed. Andrew Ortony, 19-41. Cambridge: Cambridge University Press, 1993.

Blumer, Herbert. *Symbolic Interactionism: Perspective and Method*. Englewood Cliffs, NJ: Prentice Hall, 1969.

Boden, Margaret A. *Artificial Intelligence and Natural Man*. Sussex, England: The Harvester Press, 1977.

_______. *The Philosophy of Artificial Intelligence*. New York: Oxford University Press, 1990.

Boethius, *Consolation of Philosophy*, trans. V.E.Watts. London: The Folio Society, 1998.

Bolter, David J. *Turing's Man: Western Culture in the Computer* Age. Chapel Hill: The University of North Carolina Press, 1984.

Boole, George. *The Mathematical Analysis of Logic: Being An Essay Towards a Calculus of Deductive Reasoning*. Cambridge: Cambridge University Press, 1847.

_______. *An Investigation of the Laws of Thought: On Which Are Founded The Mathematical Theories of Logic and Probability*. New York: Dover, 1854.

________. "Logic." In *George Boole: Selected Manuscripts on Logic and its Philosophy*, Ed. Ivor Grattan-Guinness and Gerard Bornett, 126-153. Boston: Birkhäuser Verlag, 1997.

________. "On the Foundations of the Mathematical Theory of Logic." In *George Boole: Selected Manuscripts on Logic and its Philosophy*, ed. Ivor Grattan-Guinness and Gerard Bornett, 63-104. Boston: Birkhäuser Verlag, 1997.

Boswell, James. *Life of Johnson*. Oxford: Oxford University Press, 1960.

Boyd, Richard. "Metaphor and Theory Change: What is "Metaphor" a Metaphor for?" In *Metaphor and Thought*, ed. Andrew Ortony, 481-532. Cambridge: Cambridge University Press, 1993.

Brewster, Edwin Tenney. *Natural Wonders Every Child Should Know*. New York: Grosset & Dunlap, 1912.

Broadbent, D. *Perception and Communication*. New York: Pergamon, 1958.

Browder, F. E. *Mathematical Developments Arising from Hilbert Problems, Proceedings of Symposia in Pure Mathematics*, vol. 28. Providence, RI: American Mathematical Society, 1976.

Brummett, Barry. *Rhetoric of Machine Aesthetics*. Westport, CT: Praeger, 1999.

Burke, Kenneth. *A Grammar of Motives*. Berkeley: University of California Press, 1969.

_____. *A Rhetoric of Motives*. Berkeley: University of California Press, 1969.

_____. *Language as Symbolic Action: Essays on Life, Literature and Method*. Berkeley: University of California Press, 1973.

______. *On Symbols and Society*, ed. Joseph R. Gusfield. Chicago: The University of Chicago Press, 1989.

Campbell, Norman. *Foundations of Science*. New York: Dover, 1957.

Carlyle, Thomas. *On Heroes and Hero-Worship, & the Heroic in History*. Berkeley: University of California Press, 1993.

Chesterton, G. K. *The Defendant*. New York: Free Port, 1972.

Chomsky, Noam. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press, 1965.

Churchland, Paul. "On the Nature of Theories: A Neurocomputational Perspective." In *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland, 251-292. Cambridge, MA: MIT Press, 1997.

_______. "Some Reductive Strategies in Cognitive Neurobiology," *Mind* XCV, 379: 279-309.

——— *Matter and Consciousness*. Cambridge, MA: Bradford Book, MIT Press, 1999.

Clark, Andy. *Being There: Putting Brain, Body and World Together Again*. Cambridge, MA: MIT Press, 1999.

_______. *Mindware: An Introduction to the Philosophy of Cognitive Science*. New York: Oxford University Press, 2001.

Coleridge, Samuel Taylor. *Specimens of the Table Talk of Samuel Taylor Coleridge*. London: Oxford University Press, 1851.

Constantinides, Helen. "The Duality of Scientific Ethos: Deep and Surface Structure," *Quarterly Journal of Speech*, 87(2001): 61-72.

Cowan, Jack D and David H. Sharp, "Neural Nets and Artificial Intelligence." In *The Artificial Intelligence Debate: False Start, Real Foundations*, ed. Stephen A. Graubard, 84-121. Cambridge, MA: MIT Press, 1988.

Crosswhite, James. *The Rhetoric of Reason: Writing and the Attractions of Argument*. Madison: The University of Wisconsin Press, 1996.

Cyphert, Dale. "Strategic Use of the Unsayable: Paradox as Argument in Gödel's Theorem," *Quarterly Journal of Speech* 84(1998): 80-93.

Dagognet, François. "Materialism: A Philosophy of Multiple Revivals," *Graduate Faculty Philosophy Journal* 22 (2000): 3-14.

Damasio, A. and H. Damasio. "Cortical Systems for Retrieval of Concrete Knowledge: The Convergence Zone Framework." In *Large Scale Neuronal theories of the Brain,* ed. C. Koch and J. Davis, 135-148. Cambridge, MA: MIT Press, 1994.

Dennett, Daniel. *Brainstorms*. Brighton, England: Harvester Press, 1981.

_______. 'Consciousness.' In *The Oxford Companion to the Mind,* ed. R. L. Gregory, 162. Oxford: Oxford University Press, 1987.

________. "True Believers: The Intentional Strategy and Why it Works." In *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland, 57-80. Cambridge, MA: MIT Press, 1997.

Descartes, René. "Meditations II, VI." In *Problems in Mind: Readings in Contemporary Philosophy of Mind*, ed. Jack S. Crumley II. 21-33. Mountain View, CA: Mayfield Press, 2000.

Diderot, Denis. *Oeuvres Complès,* trans. Roger Lewinter. Paris: Club Français du Livre, 1969-1973.

Dreyfus, Hubert L. and Stuart E. Dreyfus, *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer.* New York: Free Press, 1986.

_________. "Making a Mind Versus Modeling the Brain: Artificial Intelligence Back at a Branchpoint." In *The Artificial Intelligence Debate: False Starts, Real Foundations*, ed. Stephen R. Graubard, 15-43. Cambridge, MA: MIT Press, 1988.

Dreyfus, Hubert L. "From Micro-Worlds to Knowledge Representation: AI at an Impasse." In *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland, 143-182. Cambridge, MA: MIT Press, 1997.

Edwards, Paul N. *Computers and the Politics of Discourse in Cold War America.* Cambridge, MA: MIT Press, 1996.

Einstein, Albert. "Maxwell's Influence on the Development of the Conception of Physical Reality." In *James Clark Maxwell: A Commemoration Volume 1831-1931* (Essays by Sir J. J. Thomson, Max Planck, Albert Einstein, Sir Joseph Larmor, Sir James Jeans, William Garnett, Sir Ambrose Fleming, Sir Oliver Lodge, Sir R. T. Glazebrook, Sir Horace Lamb), Cambridge: Cambridge University Press, 1931.

Elman, J. L. "Distributed Representations, Simple Recurrent Networks, and Grammatical Structure." In *Connectionist Approaches to Language Learning*, ed. David S. Touretzky, 91-122. Dordrecht, Netherlands: Kluwer, 1991.

Emery, Jack. *Breaking the Code*. London: BBC productions, 1997.

Empedocles, 'Fragments.' In *Ancilla to the Pre-Socratic Philosophers*, ed. Kathleen Freeman, *Ancilla to the Pre-Socratic Philosophers*, 50-51. Cambridge, MA: Harvard University Press, 1996.

Feenberg, Andrew. *Critical Theory of Technology*. Oxford: Oxford University Press, 1991.

Fernandez-Duque, Diego and Mark L. Johnson, "Attention Metaphors: How Metaphors Guide the Cognitive Psychology of Attention." In *Cognitive Science* 23 (1999): 83-116.

Fields, Chris. "Measurement and Computational Description." In *Machines and Thought: The Legacy of Alan Turing, Vol. 1*, ed. P. J. R. Millican and A. Clark, 165-177. Oxford: Clarendon Press, 1996.

Fodor, Jerry A. and Zenon W. Pylyshyn, "Connectionism and Cognitive Architecture" A Critical Analysis." In *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland, 309-350. Cambridge, MA: MIT Press, 1997.

Foerster, Heinz Von. "Metaphysics of an Experimental Epistemologist." In *Brain Processes, Theories and Models: An International Conference in the Honor of W. S. McCulloch 25 years after His Death*, ed. Roberto Moreno-Diaz and José Mira-Mira, 3-10. Cambridge, MA: MIT Press, 1996.

Fowler, William A. "The Quest for the Origins of the Elements," *Science* 226 (1984): 922-935.

Fulton, John F. *Selected Readings in the History of Physiology*. Springfield, IL: Charles C. Thomas, 1930.

Galton, Anthony. *The Church-Turing Thesis: Its Nature and Status* in *Machines and Thought: The Legacy of Alan Turing, Vol. 1*, ed. P. J. R. Millican and A. Clark, 137-164. Oxford: Clarendon Press, 1996.

Gamov, G. and C. L. Critchfield. *Theory of Atomic Nucleus and Nuclear Energy-Source*. Oxford: Clarendon Press, 1949

Garver, Newton and Seung-Chong Lee, *Derrida & Wittgenstein.* Philadelphia: Temple University Press, 1994.

Gentner, Dedre and Michael Jeziorski, "From Metaphor to Analogy in Science." In *Metaphor and Thought, Second Edition,* ed. Andrew Ortony, 447-480. Cambridge: Cambridge University Press, 1993.

Gershenson, Daniel E. and Daniel A. Greenberg. *Anaxagoras and the Birth of Physics.* New York: Blaisdell, 1964.

Gödel, Kurt. "On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems I." In *Frege and Gödel: Two Fundamental Texts in Mathematical Logic,"* trans and ed., Jean Van Heijenoort, 83-108. Cambridge, MA: Harvard University Press, 1970.

Goodman, Nelson. *Ways of Worldmaking.* Indianapolis: Hackett Pub, 1978.

Graham, William. *The Creed of Science.* London: Kegan Paul, 1881.

Greene, Brian. *The Elegant Universe: SuperStrings, Hidden Dimensions and the Quest for the Ultimate Theory.* New York: Vintage Books, 2000.

Greene, John O. "Evaluative Cognitive Explanations of Communicative Phenomena." In *Quarterly Journal of Speech,* 70: 241-254.

Gross, Alan G. "On the Shoulders of Giants: Seventeenth-century Optics as an Argument Field," *Quarterly Journal of Speech* 74 (1988): 1-17.

_______. *The Rhetoric of Science.* Cambridge, MA: Harvard University Press, 1996.

Guyton, Arthur C. *Basic Neuroscience.* Philadelphia: W. B. Saunders Company, 1991.

Habermas, Jurgen, *Toward a Rational Society,* trans J. Shapiro. Boston: Beacon Press, 1970.

_______.*Knowledge and Human Interests,* trans. J. J. Shapiro. Boston: Beacon Press, 1971.

Hacking, Ian. *Representing and Intervening.* Cambridge: Cambridge University Press, 1983.

Haugeland, John. *Mind Design II: Philosophy, Psychology, Artificial Intelligence.* Cambridge, MA: MIT Press, 1997.

Heidegger, Martin. *The Question Concerning Technology and Other Essays*, trans. William Lovitt. New York: Harper & Row, 1977.

Heijenoort, J.van. *From Frege to Gödel*. Cambridge, MA: Harvard University Press, 1967.

Hobbes, Thomas. "Leviathan (Selections)." In *Classics of Modern Political Theory: Machiavelli to Mill*, ed. Steven M. Cahn, 78-196. New York: Oxford University Press, 1997.

Hodges, Andrew. *Alan Turing: The Enigma*. New York: Walker & Company, 2000.

Holton, Gerard. *The Scientific Imagination*. Cambridge, MA: Harvard University Press, 1998.

Holyoak, Keith J. and Paul Thagard. *Mental Leaps – Analogy in Creative Thought*, Cambridge, MA: MIT Press, 1995.

Hoyle, F. and R. A. Lyttleton, "A New Model for the Expanding Universe," *Monthly Notices of the Royal Astronomical Society* 108 (1948): 372-382.

Hoyle, Fred. *The Nature of the Universe*. Oxford: Blackwell, 1950.

Hume, David. *A Treatise of Human Nature*. Oxford: Clarendon Press, 1965.

Johnson-Laird, P. N. *The Computer and the Mind: An Introduction to Cognitive Science*. Cambridge, MA: Harvard University Press, 1988.

Jonsen, Albert R and Stephen Toulmin, *The Abuse of Casuistry: A History of Moral Reasoning*. Berkeley: University of California Press, 1988.

Kneebone, G. T. *Mathematical Logic and the Foundations of Mathematics*. New York: Dover, 2001.

Knoepsel, Kenneth J. "The Narrative Matter of Mathematics: John Dee's Preface to the *Elements* of Euclid of Megara," *Philosophical Quarterly* 66(1987): 27-46.

Kragh, Helge. *Cosmology and Controversy: The Historical Development of Two Theories of the Universe*. Princeton, NJ: Princeton University Press, 1996.

Kuhn, Thomas S. *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press, 1996.

Laita, L. M. "Influences on Boole's Logic: The Controversy between William Hamilton and Augustus de Morgan," *Ann. Sci* 36 (1979): 45-65.

Lakoff, George and Mark Johnson. *Metaphors We Live By*. Chicago: The University of Chicago Press, 1980.

______. "The Contemporary Theory of Metaphor." In *Metaphor and Thought*, ed. Andrew Ortony, 202-251. Cambridge: Cambridge University Press, 1993.

La Mettrie, Julien Offray de. *Machine Man and Other Writings*, trans. Ann Thomson. Cambridge: Cambridge University Press, 1966.

Leibniz, G.W. *Theodicy*, ed. Diogenes Allen. New York: Bobbs-Merrill, 1966.

______. "New System of the Nature of Substances and their Communication, and of the Union Which Exists Between the Soul and the Body." In *G. W. Leibniz: Philosophical Texts*, ed. R. S. Woolhouse and Richard Franks, 143-152. New York: Oxford University Press, 1998.

______. "Monadology." In *G. W. Leibniz: Philosophical Texts*, ed. R. S. Woolhouse and Richard Franks, 266-281. New York: Oxford University Press, 1998.

Lemaitres, Georges. "The Beginning of the World from the Point of View of Quantum Theory," *Nature* 127 (1931): 706.

Lettvin, J. Y, H. Maturana, W. S. McCulloch and W. H. Pitts, "What the Frog's Eye Tells the Frog's Brain," *Proceedings of the IRE* 47 (1959): 1940-1951.

Lewis, C. I. *A Survey of Symbolic Logic*. Berkeley: University of California Press, 1918.

Locke, John. *An Essay Concerning Human Understanding*, ed. J. W. Jolton. London: Dent, 1961.

Loeb, J. *Comparative Physiology of the Brain and Comparative Psychology*. New York: Putnam, 1900.

Loeb, J. "The Significance of Tropisms for Psychology." In *The Mechanistic Conception of Life*, ed. D. Fleming, 40-41. Cambridge, MA: Belknap Press of Harvard University, 1964.

Lyne, John and Henry F. Howe, "Punctuated Equilibria: Rhetorical Dynamics of a Scientific Controversy." In *Landmark Essays on Rhetoric of Science Case Studies*, ed. Randy Allen Harris, 65-80. Hillsdale, NJ: Lawrence Erlbaum, 1997.

Lucas, J. R. "Minds, Machines, and Gödel: A Retrospect." In *Machines and Thought: The Legacy of Alan Turing,* vol. 1, ed. Peter Millican and Andy Clark, 103-124. Oxford: Oxford University Press, 1996.

MacHale, Desmond. *George Boole: His Life and Work.* Dublin: Boole Press Limited, 1985.

Marcuse, Herbert. *One-Dimensional Man.* Boston: Beacon, 1964.

Marr, David. *Vision.* San Francisco: W. H. Freeman, 1982.

Martinich, A. P. "Thomas Hobbes." In *Modern Philosophers: From Descartes to Nietzsche*, ed. Steven M. Emmanuel, 28-42. Oxford: Blackwell, 2001.

Mayer, Richard E. "The Instructive Metaphor: Metaphoric Aids to Students' Understanding of Science." In *Metaphor and Thought,* ed. Andrew Ortony, 561-578. Cambridge: Cambridge University Press, 1993.

McCarthy, John. "Mathematical Logic in Artificial Intelligence." In *The Artificial Intelligence Debate: False Starts, Real Foundations*, ed. Stephen R. Graubard, 297-311. Cambridge, MA: The MIT Press, 1988.

_______. "Ascribing Mental Qualities to Machines." In *Formalizing Common Sense: Papers by John McCarthy,* ed. Vladimir Lifschitz, 93-118. Norwood, NJ: Ablex Publishing Corporation, 1990.

McCarthy, John and P. J. Hayes, "Some Philosophical Problems From the Standpoint of Artificial Intelligence." In *Formalizing Common Sense: Papers by John McCarthy,* ed. Vladimir Lifschitz, 21-63. Norwood, NJ: Ablex Publishing Corporation, 1990.

_________. "Programs with Common sense." In *Formalizing Common Sense: Papers by John McCarthy,* ed. Vladimir Lifschitz, 9-16. Norwood, NJ: Ablex Publishing Corporation, 1990.

McClelland, Jay L., Jerome A. Feldman, B. Adelson, Gordon H. Bower and Drew McDermot, *Connectionist Models and Cognitive Science: Goals, Directions and Implications.* Report to the National Science Foundation, June 1986.

McCulloch, Warren S. and Walter S. Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity." In *The Philosophy of Artificial Intelligence*, ed. Margaret A. Boden, 22-39. Oxford: Oxford University Press, 1990.

McFarland, David J. "Animals as Cost-Based Robots." In *The Philosophy of Artificial Life*, ed. Margaret A. Boden, 179-205. Oxford: Oxford University Press, 1996.

McKeon, Richard. *On Knowing – The Natural Sciences,* ed. David B. Owen and Zahava McKeon. Chicago: The University of Chicago Press, 1994.

McPhee, Robert D. "Cognitive Perspectives on Communication: Interpretive and Critical Responses." In *The Cognitive bases of Interpersonal Communication*, ed. Dean E. Hewes, 225-245. Hillsdale, NJ: Lawrence Erlbaum, 1995.

Merton, Robert K. "The Normative Structure of Science." In *The Sociology of Science: Theoretical and Empirical Investigations*, 267-278. Chicago, IL: The University of Chicago Press, 1973.

Michie, Donald. "Turing's Test and Conscious Thought." In *Machines and Thought: The Legacy of Alan Turing,* vol. 1, ed. Peter Millican and Andy Clark, 27-34. Oxford: Oxford University Press, 1996.

Miller, Richard B. *Casuistry and Modern Ethics.* Chicago: The University of Chicago Press, 1996.

Milton, John. *John Milton: The Complete Poems,* ed. John Leonard. London: Penguin Books, 1998.

Minsky, Marvin and Seymour Papert, *Perceptrons: An Introduction to Computational Geometry.* Cambridge, MA: MIT Press, 1969.

Minsky, Marvin. *The Society of Mind.* New York: Simon and Schuster, 1985.

_______. "Connectionist Models and their Prospects." In *Connectionist Models and Their Implications: Readings from Cognitive Science,* eds. David Waltz and Jerome A. Feldman, vii-xvi. Norwood, NJ: Ablex Publishing Corporation, 1988.

_______. "A Framework for Representing Knowledge." In *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland, 111-142. Cambridge, MA: Bradford Book, MIT Press, 1997.

Mira, J. "Reverse Neurophysiology: The "Embodiments of Mind" Revisited." In *Brain Processes, Theories and Models: An International Conference in the Honor of W.S.McCulloch 25 years after His Death* , ed. Roberto Moreno-Diaz and José Mira-Mira, 37-49. Cambridge, MA: MIT Press, 1996.

Moravec, Hans. *Mind Children.* Cambridge, MA: Harvard University Press, 1988.

Morenzo-Diaz, R. and J. Mira, "Logic and Neural Nets: Variations on Themes by W. S. McCulloch." In *Brain Processes, Theories and Models: An International*

Conference in the Honor of W. S. McCulloch 25 years after His Death, ed. Roberto Moreno-Diaz and José Mira-Mira, 24-36. Cambridge, MA: MIT Press, 1996.

Nagel, Thomas. *The View from Nowhere*. New York: Oxford University Press, 1986.

Narayanan, Ajit. "Intentional Stance and Imitation Game." In *The Legacy of Alan Turing: Machines and Thought, Vol. I*, ed. Peter Millican and Andy Clark, 63-80. Oxford: Oxford University Press, 1999.

Neumann, J. von. *Cerebral Mechanisms in Behavior-the Hixon Symposium*. New York: John Wiley, 1948.

________. "First Draft of a Report on the EDVAC." In *The Origins of Digital Computers*, ed. Brian Randell, 355-364. New York: Springer-Verlag, 1973.

Newell, Allen and Herbert A. Simon, "Computer Science as Empirical Inquiry: Symbols and Search." In *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland, 81-110. Cambridge, MA: MIT Press, 1997.

Ortony, Andrew. "Similarity in Similes and Metaphors." In *Metaphor and Thought*, ed. Andrew Ortony, 342-356. Cambridge: Cambridge University Press, 1993.

Penrose, Roger. *The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics*. New York: Penguin Books, 1989.

Perelman, CH. and L. Olbrechts-Tyteca. *The New Rhetoric: A Treatise on Argumentation.* Notre Dame, IN: The University of Notre Dame Press, 1969.

Perutz, Max. "Letter." In *Science* 164: 1537-1538.

Pitts, Walter H. and Warren S. McCulloch, "How We Know Universals, the Perception of Auditory and Visual Forms," *Bulletin of Mathematics and Biophysics* 9 (1947): 127-147.

Poincare, Henri. *Science and Method*, trans. Francis Maitland. New York: Dover Publications, 1952.

________. "Reading: Space and Geometry." In *Space from Zeno to Einstein: Classic Readings with a Contemporary Commentary*, ed. Nick Huggett, 233-235. Cambridge, MA: MIT Press, 1999.

Polanyi, Michael. *Personal Knowledge: Towards a Post-Critical Philosophy*. Chicago: The University of Chicago Press, 1974.

Poulakos, John. "Towards a Sophistic Definition of Rhetoric." In *Contemporary Rhetorical Theory*, ed. John Luis Lucaites, Celeste Michelle Condit and Sally Caudill, 26-38. New York: Guilford Press, 1999.

Prelli, Lawrence J. *A Rhetoric of Science: Inventing Scientific Discourse*. Columbia: University of South Carolina Press, 1989.

Putnam, Hilary. "Computational Psychology and Interpretation Theory." In *Artificial Intelligence*, ed. Rainer Born, 1-17. New York: St. Martin's Press, 1987.

Pylyshyn, Zenon W. "Computation and Cognition: Issues in the Foundations of Cognitive Science." In *Behavioral and Brain Sciences* 3 (1980): 111-169.

______. "Metaphorical Imprecision and the "Top-Down" Research Strategy." In *Metaphor and Thought*, ed. Andrew Ortony, 543-558. Cambridge: Cambridge University Press, 1993.

Quine, Willard Van Orman. *Word & Object*. Cambridge, MA: MIT Press, 1960.

_________. *Selected Logic Papers*. New York: Random House, 1966.

Ramsey, William; Stephen Stich and Joseph Garon, "Connectionism, Eliminativism and the Future of Folk Psychology." In *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland, 351-376. Cambridge, MA: MIT Press, 1997.

Ratner, Joseph. *Intelligence in the Modern World: John Dewey's Philosophy*. New York: Random House, 1939.

Ray, Thomas S. "An Approach to the Synthesis of Life." In *The Philosophy of Artificial Life*, ed. Margaret A. Boden, 111-145. Oxford: Oxford University Press, 1996.

Reddy, Michael J. "The Conduit Metaphor: A Case of Frame Conflict in our Language about Language." In *Metaphor and Thought*, ed. Andrew Ortony, 164-201. Cambridge: Cambridge University Press, 1993.

Reeves, Byron and Clifford Nass. *The Media Equation: How People Treat Computers, Television and New Media like Real People and Places*. Cambridge: Cambridge University Press, 1996.

Rorty, Richard. *Philosophy and the Mirror of Nature*. Princeton, NJ: Princeton University Press, 1979.

Rosenberg, Jay F. "Connectionism and Cognition." In *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland, 293-308. Cambridge, MA: MIT Press, 1997.

Rosenblatt, F. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain," *Psychological Review* 65, no. 6 (1958): 386-408.

________. "Strategic Approaches to the Study of Brain Models." In *Principles of Self-Organization*, ed. H. von Foerster, 75-85. Elmsford, NY: Pergamon Press, 1962.

________.*Principles of Neurodynamics*. New York: Spartan, 1962.

Rumelhart, D. E, G. E. Hinton and R. J. Williams, "Learning Internal Representations by Error Propagation." In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. I: Foundations*, eds. David E. Rumelhart, James L. McClelland and the PDP Research Group, 318-362. Cambridge, MA: MIT Press, 1986.

Rumelhart, David E et al. *Parallel Distributed Processing: Exploration in the Microstructure of Cognition, Vol I.* Cambridge, MA: Bradford Book, MIT Press 1986.

_______. *Parallel Distributed Processing: Exploration in the Microstructure of Cognition, Vol II.* Cambridge, MA: Bradford Book, MIT Press, 1986.

Rumelhart, Davide E. "Some Problems with the Notion of Literal Meanings." In *Metaphor and Thought*, ed. Andrew Ortony, 71-82. Cambridge: Cambridge University Press, 1993.

________. "The Architecture of Mind: A Connectionist Approach." In *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland, 205-232. Cambridge, MA: Bradford Book, MIT Press, 1997.

Russell, Bertrand. "On Scientific Method in Philosophy." In *Mysticism and Logic*. New York: Norton & Company, 1929.

________. *The Problems of Philosophy.* Oxford: Oxford University Press, 1959.

Sayre, K. M. "Cybernetics." In *Philosophy of Science, Logic and Mathematics in the 20th century*, ed. Stuart G. Shanker, 292-314. New York: Routledge, 1996.

Scott, Robert L. "On Viewing Rhetoric as Epistemic," *Central States Speech Journal* 27 (1967): 258-266.

Searle, John R. "Minds, Brains and Programs." In *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland, 183-204. Cambridge, MA: MIT Press, 1997.

Sejnowski, T and C. Rosenberg, "Parallel Networks that Learn to Pronounce English Text," *Complex Systems* 1 (1987): 145-168.

Shanker, Stuart G. "Introduction." In *Philosophy of Science, Logic and Mathematics in the 20th century*, Routledge History of Philosophy, Volume ix, ed. Stuart G. Shanker, 1-24. Routledge: London, 1996.

______. "Descartes' Legacy: The Mechanist/Vitalist Debates." In *Philosophy of Science, Logic and Mathematics in the 20th century*, vol ix, ed. Stuart G. Shanker, 345-360. London: Routledge, 1996.

Shannon, Claude E. *A Symbolic Analysis of Relay and Switching Circuits*, Master's Thesis, Massachusetts Institute of Technology; published in *Transactions of the American Institute* 57 (1938): 1-11.

Shannon, Claude E. and Warren Weaver. *The Mathematical Theory of Communication.* Urbana: The University of Illinois Press, 1962.

Simon, Herbert A. "Thinking by Computers." In *Mind and Cosmos: Essays in Contemporary Science and Philosophy*, ed. Robert. G. Colodny, 15-25. Pittsburgh: The University of Pittsburgh Press, 1966.

________. *The Sciences of the Artificial.* Cambridge, Massachusetts: MIT Press, 1969.

________. "Machine as Mind." In *Machines and Thought: The Legacy of Alan Turing Vol I*, ed. P. J. R. Millican and A. Clark, 90-96. Oxford: Clarendon Press, 1996.

Skinner, B. F. *Beyond Freedom and Dignity*. New York: Knopf, 1971.

______. "Excerpts from About Behaviorism." In *Problems in Mind: Readings in Contemporary Philosophy of Mind*, ed. Jack S. Crumley II, 59-67. Mountain View, California: Mayfield Publishing Company, 2000.

Sloman, Aaron. "Beyond Turing Equivalence." In *Machines and Thought: The Legacy of Alan Turing*, vol. 1, ed. Peter Millican and Andy Clark, 179-219. Oxford: Oxford University Press, 1996.

Smith, Crosbie and M. Norton Wise, *Energy and Empire: A Biographical Study of Lord Kelvin.* Cambridge: Cambridge University Press, 1989.

Smolensky, Paul. "Connectionist Modeling: Neural Computation/Mental Connections." In *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed. John Haugeland, 233-250. Cambridge, Massachusetts: Bradford Book, MIT Press, 1997.

Sternberg, Robert J. *Metaphors of Mind: Conceptions of the Nature of Intelligence* Cambridge: Cambridge University Press, 1990.

Taylor, Charles Alan. *Defining Science: A Rhetoric of Demarcation*. Madison: The University of Wisconsin Press, 1996.

Thomson, Ann. "La Mettrie, Machines, and the Denial of Liberty," *Graduate Faculty Philosophy Journal* 22 (2000): 71-86.

Thorndike, Edward Lee. *Human Learning*. New York: Appleton-Century Crofts, 1931.

________. "Why Study Animal Psychology?" In *Comparative Psychology*, ed. F. A. Moss, 1-6. New York: Prentice-Hall, 1942.

__________. *Selected Writings From a Connectionist's Psychology*. New York: Appleton-Century-Crofts, 1949.

Tunbridge, Paul. *Lord Kelvin, His Influence on Electrical Measurements and Units*. London: Peter Peregrinus Ltd on behalf of the Institution of Electrical Engineers, 1992.

Turing, A. M. "Lecture to the London Mathematical Society, 20 February, 1947." In *A.M.Turing's Ace Report of 1946 and other papers,* " Vol.10, Charles Babbage Institute Reprint Series for the History of Computing, 106-124.Cambridge, Massachusetts: MIT Press, 1986.

________. "On Computable Numbers, With an Application to the *Entscheidungsproblem*." *Proceedings London Mathematical Society,* 43, 544-546.

______. "Computing Machinery and Intelligence." In *Mind Design II*, ed. John Haugeland, 29-56. Cambridge, Massachusetts: MIT Press, 1997.

Turkle, Sherry. *The Second Self: Computers and the Human Spirit.* New York: Simon and Schuster, 1984.

Ukrainian Chess Online, "Kasparov-Deep Blue, 1997," http://chess-sector.odessa.ua/kas_deep.htm, 1997.

Weaver, Richard. *The Ethics of Rhetoric*. La Salle, IN: Gateway, 1953.

______. *Language is Sermonic*. ed. Richard L. Johannesen, Rennard Strickland and Ralph T. Eubanks. Baton Rouge: Louisiana State University Press, 1970.

Weir, Michael K. "Putting the Mind Inside the Head." In *Brain Processes, Theories and Models: An International Conference in the Honor of W.S.McCulloch 25 years after His Death* , ed. Roberto Moreno-Diaz and José Mira-Mira, 54-63. Cambridge, MA: MIT Press, 1996.

Weizenbaum, Joseph. *Computer Power and Human Reason: From Judgment to Calculation.* New York: W.H.Freeman and Company, 1976.

Whately, Richard. *Elements of Logic*. Boston: James Munroe and Company, 1859.

Wiener, Norbert. *Cybernetics, or Control and Communication in the Animal and the Machine*. New York: Wiley, 1948.

Willard, Charles Arthur. *Argumentation and the Social Grounds of Knowledge*. Tuscaloosa: The University of Alabama Press, 1983.

Willshaw, David. "Non-Symbolic Approaches to Artificial Intelligence and the Mind," *Philosophical Transactions: Physical Sciences and Engineering* 349 (1994): 87-101.

Wilson, Margaret Dauler. "Leibniz and Materialism." In *Ideas and Mechanism: Essays on Modern Philosophy*, ed. Margaret Dauler Wilson, 388-406. Princeton, NJ: Princeton University Press, 1999.

Wilson, Margaret Dauler. *Ideas and Mechanism: Essays on Modern Philosophy*. Princeton, NJ: Princeton University Press, 1999.

Wittgenstein, Ludwig. *Philosophical Investigations*. trans. G.E.M. Anscombe. Oxford: Basil Blackwell, 1958.

______. *Wittgenstein Lectures: Cambridge 1930-1932*, ed. D. Lee. Oxford: Blackwell, 1980.

________. *Culture and Value*, trans. P. Winch. Oxford: Basil Blackwell, 1980.

Woolgar, Steve. "Representation, Cognition and Self: What Hope for an Integration of Psychology and Sociology?" In *Ecologies of Knowledge: Work and Politics in Science and Technology*, ed. Susan Leigh Star, 155-179. Albany: State University of New York Press, 1995.

VITA

Roy Joseph is an instructor in the Department of Communication and Rhetorical studies at Duquesne University, Pittsburgh. Mr. Joseph graduated in August of 1998 with a Master of Arts in Communication Studies from the University of Nevada-Las Vegas. His other degrees include a Master of Philosophy in English, a Master of Arts in English and Bachelor of Arts in English in the years 1995, 1994 and 1992 respectively from Pondicherry University, India. After completing his Doctorate of Philosophy in Speech Communication from Texas A&M University, Mr. Joseph will begin his tenure track position as Assistant Professor in Duquesne University. Mr. Joseph resides at 603 Nash Street, Pittsburgh, Pennsylvania, 15212.