12-2012

Math, Minds, Machines

Christopher V. Carlile
University of Tennessee - Knoxville, ccarlil1@utk.edu

Math, Minds, Machines

Christopher Carlile
University of Tennessee, Knoxville

December 2, 2012

Dear reader,

While this paper is ostensibly a thesis on human intelligence and the possibility of artificial intelligence, it *is* much more than that to me. Its exigence is untraceable to any specific interaction, experience, or even deadline, and I'd like to provide some of that context for you. If you honestly don't care, go ahead and skip this letter. I won't feel bad.

First and foremost this paper is an integration of two topics which I am pretty sure I spent more time as an undergraduate thinking about than anything else. The first is mathematics, and by virtue of being an honors math student I was spending untold amounts of time doing homework, studying for tests, and the like. While I have decided, only really in the last year, that going to graduate school for math was not for me, I am very pleased with my choice of major. Given the chance to do it all over again, I would only change by taking more math. It equipped me with the arsenal of deductive abilities which helped guide my interests and inquiries, and now I feel i would have been wasting my time doing almost anything else. I am especially grateful to Drs. Plaut, Thistlethwaite, and Nicoara for being outstanding and patient advisors, mentors, and most importantly teachers.

The second topic is consciousness, and it's what I was thinking about on weekends and bank holidays when my math responsibilities were done (and sometimes when said responsibilities were not done, natch). A smattering of psychology and philosophy classes and reading throughout my academic career helped guide this interest, but two sources in particular deserve mention. The first is Douglas Hofstadter's classic work on the topic, *Gödel, Escher, Bach*. I read it when I was first beginning as a math student, and I blame my rapid increase in interest in consciousness in large part on the concurrence of the two. I drew on it heavily for inspiration, and if I could give Dr. Hofstadter a high-five then I totally would, but I'll settle for giving him mention in this letter (and bibliography where appropriate, of course). The other source for this inspiration comes from Dr. Neil Greenberg, whose freshman honors seminar I stayed conscious in enough to realize I didn't get the full message of the class, and the term paper I submitted for the full Art & Organism seminar in spring 2012 forms much of the first section of this paper. A&O quickly became one of the best classes I had taken, and the topics treated in the class are how much of this paper came to be written.

Further shout-outs to the Chancellor's Honors program for the numerous of avenues of support they offer and for making me study abroad, and for my friends, family, and mentors who stimulated me and encouraged me to get this thing done. You know who you are, and generating an exhaustive list would be tedious.

Finally, this thesis is an amalgamation of things I learned as an undergraduate which piqued my interest. It is a creative and meandering work, and often times I reject a more academic tone in favor of my own creative voice. I know it is not a thesis in the traditional math honors sense, especially since it contains nothing in the way of my own mathematical results, but much of what I say are my own deductions, observations, and hypothesis. It's better this way, I think, that you know this at the outset of the paper, and by reading it as such I believe the experience is closer to what I had intended. If you learn something new, even if it's about slime mold, then I will consider this paper a success. If I made you think after you put the paper down, then that's even better but I'm not holding out for it.

Thanks,
Christopher Carlile

Submitted as a requirement for the degree of Bachelor of Science in Mathematics, Honors concentration, and as a requirement for the Chancellor's Honors Program at the University of Tennessee, Knoxville

1 Introduction

Life on planet Earth is hard. The drama is the same for all creatures: around every corner there are predators to outrun, scarce food to be eaten, and even fellow members of the species to fight over hunting grounds and potential mates. For the last few tens of thousands of years, however, humanity has increasingly distinguished itself by rising above that natural order and harnessing its environment to solve these problems. Driving this trend is the ever-increasing intelligence of humanity - our ability to *know* things about the environment, and harness that knowledge, is what gave us our sedentary, agrarian lifestyle, ability to irrigate, and in more recent years enabled us to harness the electron to power a variety of machines that make a multitude of forms of human labor obsolete. It has increased our food supply, let us live twice as long as our primitive ancestors, and spread all over the planet.

Yet overcoming our place in the natural order seems only halfway complete. Yes, several diseases that have plagued humanity have become extinct at our own hand. We have unparalleled access to information which allows us to make better decisions and gain a fuller picture of the world around us. Still we work for our food, though not directly as once was the case. Major media outlets would have us believe we are constantly under threat from outside forces, whether it's the plague of the week or terrorist forces in faraway parts of the world. To top it all off competition is alive and well - for food, for water, the next promotion or the cute girl at the end of the bar. Our human condition has remained essentially unchanged, for all the superficial advances our big brains have afforded us.

This paper is, at its core, an attempt to describe a *post*human condition, a new paradigm of doing business and allocating resources that exists outside of the biological battle that has raged for billions of years. The key to this is realizing a machine that can think better than we can, an Artificial Intelligence that is insusceptible to the biases and errors of human thinking. With the ability to "know" more than is feasible to represent in the human mind, and to reason faster and with fewer errors, such an AI would be the last nail in the coffin for our natural shortcomings. To that end, this paper is divided into three sections. The first describes the history, purpose, and use of mathematics. In addition to being a subject very near to my heart, math forms the foundation of all human achievements, whether directly realized by math or not. The abstraction of reason allows us to make statements independent of the natural world, but thanks to the natural world's arising from deterministic forces, it is the most powerful tool for understanding our world. I treat a variety of topics, particularly from the 20th century but drawing from all over the history of mathematics, that have caught my interest, some proofs I particularly like, and important developments in the closely related field of computer science, which according to Dr. MacLennan of the CS department at UT is "mathematics in the imperative mode." Not all of it is necessary for the discussion on AI, but as AI essentially arises from mathematics, it's a perfectly good place to start.

The second part will deal with the much slipperier (figuratively and literally) human brain. The human brain and all of its associated methods of computation are, for lack of a better model, what our AI will resemble when it is fully matured. I'll examine the role of language in knowledge, give a quick overview of the nervous system, and talk about some other topics regarding the epistemological nature of human knowledge, including the map-territory relation as well as the process of evolution and how it contributes to the human condition. The final section will integrate the first two, and will be an overview of the field of AI as a whole, including a history of past approaches, and overview of current ones, and a look to the future of what AI will be. Examples and tangents abound, and in the spirit of free association it's included if I've found it to be particularly interesting.

2 A Brief History of Math, Euclid, and Cantor

Inquiry and expression are nothing new to the human race. From the first human to skin an animal to warm himself in the winter to the boom of the agricultural revolution, knowledge about our environment, and the resulting ability to harness it, has increased our chances of survival dramatically, propelled us to the top of the food chain and spread us all over the world. It is only relatively recently that math and science as we know it (namely, math since the introduction of set theory and science and the adoption of the scientific method) came about, but man has always felt a desire to know more about nature and seek out truth in the world.

What is truth, then? As far as I know, it's still wide open for debate, as the amount of information realized in a person's consciousness is just a small fraction of the information that reaches the brain, which is itself an even smaller fraction of the amount of information capable of being absorbed. This means that our experience and reason are all we have to understand the universe, and we use language to corroborate our findings with our fellow humans. Expression is the means by which we arrive at common ground with regards to to what truth actually is, or more precisely what truth is not. If two people disagree, then at least one must be wrong, but in the case that people agree we can only say that an experience that would evoke a disagreement has not been had by either party. Language enables us to verify truth, but only insofar as to what degree the consensus approximates that truth.

Math is an attempt to solve this problem. By minimizing that which relies on consensus, and establishing strict rules of inference, disagreement in mathematics tends to the objective. Add to that the fact that every statement is either true or false with regards to its axioms, and results in mathematics are held as indisputable. As Einstein said, "as far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality." The results we obtain from mathematics all come from atomic elements which humans come up with - definitions, axioms and basic rules of deduction - and it is the quality with which they are selected that determine the efficacy of our results as being good approximations of truth.

It seems that mathematics is a purely human trait, like language. That is not to say the ideas with which mathematicians concern themselves are outside the concern of animals. Capuchin monkeys have demonstrated quantitative abilities such as counting and basic arithmetic (Beran) , but interest in the intrinsic properties of form and quantity is not only human but is also a relatively recent development. Our earliest inquiries into quantity arose as would any linguistic attempt to convey information about the world - we came up with words for concepts that we wanted to discuss with others. The need to express quantities larger than how many animals you brought home for dinner came later, and existence of languages whose only quantities are "one," "two," and "many" in extant hunter-gatherer societies (Boyer) corroborates that our original conceptions of "numbers" came as a result of necessity.

With the rise of civilization came free time for humans to spend learning more about the world around them. The passing of days, seasons and years influenced early agricultural societies, and record keeping for these cycles and for quantifying harvest was enabled after the development of writing and corresponding written representations of quantity. The earliest of these was of course the humble tally mark, the earliest written representation of "unit" and ancestor to our number one. Just like spoken language, written number systems with different symbols representing different quantities came about when the quantities in question became too large to be practically represented with tally marks. The earliest known civilization to develop such a system were the Babylonians, and by 3000BC they had a metrological system for tacking volumes, areas, and quantities. A base 60 system was chosen for its large number of divisors, and though the Hindu-Arabic decimal system is the dominant number system today, the Babylonian sexigesimal system lives on in our timekeeping and angle measurement.

The first people to explore mathematics for its own sake were the ancient Greeks, however. While earlier mathematical endeavors had been centered around problems like measurement, the Greeks went a step further by axiomatizing geometry and realizing that the natural world operated "not as the products of random or arbitrary influences, but regular and governed by determinable sequences of cause and effect" (Lloyd). It is this realization that led to the Greeks utilizing mathematical rigor and deductive reasoning, opposed to the inductive methods used by previous cultures (Heath). This turning point marked a change in mathematics from a tool to assist us in our calculations and machinations to a study where the process of arriving at results is as important as the results themselves. This new approach is critical in understanding *why* the natural world operates the way it does, instead of simply observing the resulting phenomena.

Lots of big names in math come to us from classical Greece: Pythagoras explored the then-taboo irrational numbers and Archimedes was known for streaking and specific gravity, but perhaps most influential was Euclid, not for his discoveries but for assembling the first real math textbook, his *Elements*, which was primarialy focused on geometry but also included disgused results in number theory and algebra. Key to Euclidean geometry is the parallel postulate - that given a line L and a point p, there is exactly one line passing though p and parallel to L. First introduced by Euclid in his *Elements*, it stood apart from the other four postulates on which his results are based. Postulates like the congruency of all right angles and the existence of straight line segments connecting points together seem obvious and are integral to the whole process of doing geometry, but the parallel line postulate takes a slight leap of faith to believe. Seeing as how leaps of faith generally do not hold well in mathematics, since its inception as an axiom people have been trying to prove or disprove it. This notion incorporates an idea of "flatness," or rather an absence of curvature, and since all of human experience has corroborated this idea until recently, theories and results which violated the fifth postulate were not regarded as legitimate. Even Gauss neglected to publish his work on geometry in which there exist infinitely many lines parallel to any given line and passing through a single point, a curved geometry later known as hyperbolic geometry. And it wasn't until Einstein and relativity that our current view of the universe shifted to a curved one - in relativity, mass and energy bend spacetime, meaning that what were long regarded as absolute truths in the universe are in fact simply assumptions made after humans observed and thought about what is truly only a tiny fraction of existence. The mathematics of Euclidean space are still very much applicable to our everyday purposes, but incorporated into our assumption of flatness is the knowledge that to some extent our calculations are wrong (Misner et all 43).

The arrival of set theory in the late 19^{th} century marks the rise of mathematics as we know it today. Developed by Georg Cantor and Richard Dedekind, set theory takes the formally logical notions of truth and not truth and frames it as a matter of membership in a given set (Johnson). The English sentence "the square root of two cannot be written as a fraction of two whole numbers" is translated into the set-theoretic statement "$\sqrt{2} \notin Q$," where Q is the *set* of all rational numbers - numbers written that can be expressed as a ratio of two integers. The phrasing of "expressing" a number is a canonical one. Since any number in a number system exists as soon as a definition for that number system is conceived, we can say precisely which number we are referring to in many different ways. Four doesn't exist because it is two plus two, nor because the traditional axioms of the natural numbers (formulated by Giuseppe Peano, after whom they are named) dictate that one is a natural number and every natural number has a natural number exactly one larger than it, but our system gives us a conception of "four-ness" we can all agree on. Words in natural languages don't fare as well with abstractions like "justice" or "beauty." By establishing the language of ideas like "equality" in a rigorous and formal way, however, math avoids these problems and unambiguously allows us to determine truth or falsehood.

One way of interpreting the idea of "size" was offered by Cantor in his conception of set theory. The notion of cardinality interprets size as the number of elements in a set. We say two sets are of equal size if a one-to-one correspondence between the two sets, called a bijection, can be established. His most famous result was that the natural numbers $\{1, 2, 3, \ldots\}$ have a smaller cardinality of the real numbers, the set of all numbers that can be expressed with any number of decimal points. The proof is rather slick - Cantor supposed the opposite to arrive at a proof by contradiction (fancy Latin name: *reducto ad absurdum*), and lists the real numbers. He then constructs a new real number, with a first digit different from the first digit in real number number 1, a different second digit than was found in real number number 2, and so forth, creating an unlisted real number - this approach is now called diagonalization. This was indeed shocking news to just about everyone (Dauben). Philosophers, priests, and mathematicians had long been arguing about the nature of the infinite, a concept far too large for human understanding, and along comes Cantor saying that not only is there not one true infinity, but an infinite amount of infinities. Even some of his peers - mathematicians, ostensibly driven by pure reason - dogmatically denied his result and his theory of transfinite cardinality, belittling him to the point of depression. Cantor spent the rest of his life in and out of sanatoria, dying broke and alone in a mental institution. The tantalizing attractiveness of math, its infallibility, shows its ugly side when results that challenge the way we interpret reality on a fundamental level are met with near-religious zeal in defense of the status quo. Though at the time revolutionary, his results are now accepted widely enough to consider the uncountability of the reals an elementary truth.

3 Formal Languages, Turing, Gödel

All our explorations into science are based upon the belief that our universe behaves according to predictable laws. Indeed this is the case, as the point of science is not only to explain but to predict, and no scientific theory is worth diddly if it cannot offer predictions on the matter at hand. Therefore a crucial task in any field of science is to fully understand the atomic elements of that field and how they interact. Completely understanding the way DNA expresses itself as a phenotype would unlock the key to many aspects of development and the cure for many diseases. The big problem in physics is discovering the fundamental "particles" responsible for matter, energy, and their interactions. A complete reductionist understanding with appropriately accurate and powerful simulation modeling would allow us to deduce and predict the outcome of any conceivable event.

A system that makes such statements (theorems, in the language of formal systems) deduced from some certain atomic facts is called a formal system, and mathematics is the study of such systems. There are three key features of such systems, and the first is the language of the formal system. Called a formal language, it shares some features with natural languages, like having an "alphabet" of symbols from which strings of symbols, or words, are generated (Britannica). We are limited in English as to which strings of symbols from our Latin alphabet form valid words - lexically, in the sense that creating a word without meaning is not a valid word, and phonologically, in the sense that the symbol-string "uxvcdz" is unpronounceable, and words are by definition things we say. These rules have some wiggle room, and recalling Will Ferrell's performance as James Lipton on SNL's spoof of *Inside the Actor's Studio*, his word "scrumtrulescent" has a morpheme structure reminiscent of positively-connoted words, so one can get a good idea of what Ferrell means. This understanding is predicated on an English-speaker's background and subjective experience, however, and in a formal language such rules of creating valid (or in the language of formal language *well-formed*) members of the language are objective, meaning definitively and objectively that either a string is well-formed or it is not.

The alphabet of first-order mathematical logic includes parentheses for ordering operations, symbols for denoting logical relations ($\rightarrow$, $\wedge$, etc.), the quantifiers $\forall$ and $\exists$, and a set of variables x_1, x_2, etc. While there are specific rules of formation, the formal language of first order logic borrows from natural language the readability of logical statements. $\forall x$ (Even(x) $\leftrightarrow$ $\exists y$ Equal(2 $\times$ y, x)) is read "for all x, x is even if there is a y such that 2y=x," where Even(x) and Equal(y,x) are terms that have first-order expansions. Programming languages are also notorious examples of formal language; the alphabet is the vocabulary of the programming language, code free from syntax errors is well-formed. Furthermore, the interpreter or compiler and the operating system that is running on is *also* encoded in a formal language, manipulated by the physical pathways of the hardware it is running on and representing each moment of the process of computation as the state of all the memory, input, and registers in the computer.

Beyond the language, the other features of a formal system are the starting set of atomic facts, or axioms, and rules of inference which take well-formed statements in the language and return new statements with the same truth value. The axioms, which are themselves well-formed statements, together with the rules of inference, generate the space of theorems in the formal language, which can also be considered the space of "true" statements. Normally mathematicians consider a statement then try to see if it is true by formulating a proof, starting with facts that are already known to be true and attempting to construct a series of valid inferences that lead to the statement in question. An equally valid way to obtain true statements is to start with the axioms and apply the rules of inference to every valid combination of axioms and generate a set of theorems, then apply the rules of inference to every true statement now known (that is, the axioms and all the theorems generated from them) and so on until the statement we were originally considering gets proved (Hofstadter). This is significantly less fun and practical than doing math the "old-fashioned" way.

This method generates the space of theorems, and any well-formed statement that is not generated in this manner is said to be false. However, it is possible to select axioms such that both a statement and its negation can be generated. This is undesirable as negation is an operation on a formula in the formal language, much like a rule of inference, that instead generates a new formula with an opposite truth value. This contradiction is undesirable, and if axioms are picked in such a way that no such statement appears in the theorems then the system is said to be consistent. Formally, a system is consistent if given a statement of the system, p, p AND $\neg p$ is false - either p or $\neg p$ is not a theorem. A related concept is completeness, the ability for a system to prove every statement or its negation. Instead of p AND $\neg p$ is false, we would assert p OR $\neg p$ is true (Hofstadter).

Having a formal system that is complete and coherent is obviously desirable - everything provable is true and everything true is provable means we have an infallible, powerful system for discovering truth. Indeed this was the aspiration of Bertrand Russell and Alfred Whitehead in the early 20th century with their monumental *Principia Mathematica*. By carefully selecting the proper axioms and rules of inference, they hoped to create a system for proving all mathematical truths. Defining precisely their formal system, they began from first principles and established everything rigorously, including an oft quoted proof whose punchline reads, "from this proposition it will follow, when arithmetical addition has been defined, that 1+1=2." (Stanford Encyclopedia of Philosphy)

While *Principia Mathematica* did influence contemporary thought in mathematics and logic with regards to the power of formal systems, and it did have great success proving theorems from all over mathematics, it did not become the end-all be-all of mathematical knowledge. The creators neglected to do a volume on geometry due to psychological drain inflicted on them from the first several volumes, and actually demonstrating every possible mathematical truth by producing them on paper is an impossible dream. The nail in the coffin for *Principia Mathematica* to complete its goal came in the form of Kurt Gödel and his incompleteness theorem (Stanford Encyclopedia of Philosophy).

Actually, there are two. The first states "Any effectively generated theory capable of expressing elementary arithmetic cannot be both consistent and com-

$*54{\cdot}43$. $\vdash :. \alpha, \beta \in 1 \,.\supset : \alpha \cap \beta = \Lambda \,.\equiv.\, \alpha \cup \beta \in 2$

Dem.

$$\vdash . *54{\cdot}26 \,.\supset \vdash :. \alpha = \iota\text{‘}x \,.\, \beta = \iota\text{‘}y \,.\supset : \alpha \cup \beta \in 2 \,.\equiv.\, x \neq y \,.$$

$$[*51{\cdot}231] \qquad \equiv .\, \iota\text{‘}x \cap \iota\text{‘}y = \Lambda \,.$$

$$[*13{\cdot}12] \qquad \equiv .\, \alpha \cap \beta = \Lambda \qquad (1)$$

$$\vdash . (1) . *11{\cdot}11{\cdot}35 \,.\supset$$

$$\vdash :. (\exists x, y) \,.\, \alpha = \iota\text{‘}x \,.\, \beta = \iota\text{‘}y \,.\supset : \alpha \cup \beta \in 2 \,.\equiv.\, \alpha \cap \beta = \Lambda \qquad (2)$$

$$\vdash . (2) . *11{\cdot}54 . *52{\cdot}1 \,.\supset \vdash . \text{Prop}$$

From this proposition it will follow, when arithmetical addition has been defined, that $1 + 1 = 2$.

Figure 1: Proof that 1+1=2

plete.” (Kleene) In other words, there is a statement p for which neither p nor $\neg p$ is provable. That is to say, for example, if p is false then $\neg p$ must be true, but $\neg p$ is not provable means $\neg p$ is not a theorem, though it is true. The proof is, by all means, a genius proof which exploits an ability of formal systems that can make statements about arithmetic to make arithmetically-flavored statements about the system itself, constructing a statement reminiscent of the liar’s paradox, the statement “this statement is a lie.” By establishing a correspondence between symbols in the alphabet of the formal system of *Principia Mathematica* and strings of numbers, statements (or strings of symbols) could be translated into an equivalent numerical coding, called the Gödel colding (Hofstadter). Then relationships between statements linked by a rule of inference (a step in a proof) can be encoded in a related mathematical function, and by extension provability can be encoded in a numerical way. So by encoding a the Gödel number of a statement that asserts unprovability, Gödel could find a theorem which referred to the Gödel number of an unprovable theorem, which was the theorem he found itself! And not only that - there are infinitely many such statements of unprovable and even unknowable validity.

The implications for this result were profound. In addition to shaking mathematics to the core, it raised questions about the nature of human intelligence. In a reductionist view, at each instant a neuron is either firing or not, and we can consider the whole brain, together with its rules of deduction (neural pathways) and well-formed brain states (states it is physically conceivable to exist in) as a formal system (Hofstadter). Keeping with the understanding that states of consciousness are emergent properties of the state of our brains, and knowledge a result of that, incompleteness could mean that there may be a whole host of “knowable” things that aren’t expressible as patterns of firing neurons. Of course I hesitate to use the word “knowable,” since we tend to identify truth and knowledge, but the intertwining of our perception of truth, even unknowable truth, and what our brain says we know is so deep seated that human language has developed around that concept. This all hinges on the identification of the brain as a formal system - specifically, a system that models computation called a *Turing machine*.

It is important to note that we are not identifying the brain strictly as a Turing machine for two reasons. The first is that the brain, at an initial glance, absolutely does not resemble a turing machine in either form or function. A Turing machine is an idealized computer that consists of an infinitely long *tape* divided into cells, one after the other, which is “read” by a *tape head*. The tape head refers to a finite *rules table* that contains, for each of its *state values*, instructions that modify the contents of the cell (or not and leave the contents as-is), direct the tape head to the cell to the left or the right (or not and stay on the current cell), and change the state of the machine (or not). Despite being rather straightforward, the Turing machine models what actually goes on in a computer fairly well and provided a model for formalizing the idea of a “calculable” function, effectively one that can be solved with an algorithm (Turing). To say that the brain operates in this fashion would require some carefully constructed analogies mapping the brain states, parts of the brain, and their constituent neurology onto these formally defined concepts like the tape, and until we know more about the brain or more about conscious experience in general, making this call is murky.

The other reason that we do not identify “the brain is a formal system” with “the brain is a turing machine” is the result of Alan Turing’s landmark 1936 paper “On Computable Numbers, with Application to the Entscheidungsproblem.” The exigence of this paper is due to David Hilbert, 19^{th} century mathematical heavyweight who in 1900 posed a set of 23 problems, which to him at the time were the greatest unsolved problems of mathematics (Hilbert). Included on the list was the consistency of arithmetic, which was shown by Gödel to be impossible within arithmetic, and the Reimman hypothesis, whose solution would yield insights into the structure of prime numbers and arithmetic itself yet remains notoriously unsolved. Continuing with his tradition of leadership within the mathematical community, he proposed an additional three problems in 1928, the most famous of which is the Entscheidungsproblem, or the “decision problem.” It asks whether there exists an algorithm that takes as its input a statement of first-order logic, the standard logical framework of mathematics, and returns a “yes” or “no” if the statement can be proven as a result of the logical axioms (Hodges).

Coming up with the Turing machine allowed him a definition of “algorithm” by abstracting a computer to its essential function of reading and manipulating symbols. Since each machine can be defined by its state register, its rules table, and initial tape state, all of which are finitely determined, Turing in much the same way as Gödel found a way to encode any machine in a finite string which could be fed into another machine, the universal machine, and the result it computed would be identical to the one the computed by the original Turing machine (Turing). Gödel made statements about the system of arithmetic by arithmetizing those statements, and Turing made statements about computability within a computer. The decision problem, rephrased in the language of Turing machines, is "given a description of a computer program (Turing machine), does the program come to a halt or run forever?" I will call such Turing machines that halt as "finite," Turing used the word "circle-free." A sketch of Turing's proof is as follows: suppose such a machine existed, call it D. Since D checks if a given description number (DN) represents a finite turing machine (FTM), we can set it to check for each integer n if it is the DN of a finite turing machine. At each finite step, the decision process is finite, and since each number represents at most one FTM, D is "circle free."

So for any given n, some number of integers less than or equal to n are DNs of FTMs. Call it $R(n)$. For all n, $R(n+1) = R(n)$ if n is not a FTM, and $R(n+1) = R(n) + 1$ if it is. D has a description number which we will call k. When our program D hits k, since D is finite by construction D would return that it is finite. But this would involve processing the $R(k-1)+1 = R(k)$ figures of the sequence, for which the first $R(k-1)$ figures would pose no problem but calculating the $R(k)^{th}$ amounts to "calculate the first $R(k)^{th}$ figures of H and return the $R(k)^{th}$," hence the program D gets hung up trying to compute itself. So D is *not* finite, and by contradiction no such D can exist. The parallels between Gödel's ideas and Turing's are easy enough to see, by having a system powerful enough to make statements about themselves, they crafted members of that system that referred to their own incompatibility within the system. This ability for a system to make statements about itself is known as primitive recursion, and it is crucial for these proofs that the system possesses it (Hofstadter).

By the time the paper had gotten around to being published, another paper by mathematician Alfonso Church also asserted the undecidability of the decision problem, modeling algorithms with a system of recursive functions called λ-calculus. Turing immediately saw that the problems computable in one framework could be computed in the other, so he included a proof of this equivalence in his paper. Their combined results, along with Turing's proof of their equivalence and the equivalence of subsequent models of computation, gave rise to the Church-Turing thesis that asserts "a function is computable if and only if it is computable by a Turing machine." It is this result that motivates our efforts to establish an equivalence between the brain and a Turing machine. If it is, we face the unpleasant reality of having our thoughts and actions predetermined by a set of rules and state registers, while the opposite conclusion could cast doubts on the viability of AI with a conscious experience comparable to ours. Of the two, the latter is certainly preferable, since questions of free will and mind-body dualism are really, at the end of the day, irrelevant to our everyday conscious experience. I also prefer the second because I believe the possibility of such a strong AI to be too lucrative to not try and pursue. I sincerely hope that this does not become too cliche as the paper goes on, but determining the extent of those two statements have with regards to mind requires much more research into mind itself- both from neurological and silicon-based perspectives.

4 Emergence, Recursion, and Fractals

When poets and dreamers talk about the beauty of the natural world, more often than not they talk about beautiful vistas or serene ponds, or perhaps they take delight in watching animals do their thing, operating on whatever natural instincts with which we are endowed. What they often do not talk about in poetry, however, is the atoms and molecules that make up the mountains and animals which they find so beautiful, or the electromagnetic forces that hold them all together. It seems reductionism is attractive only to scientists, who take pleasure in delving deeper than what lays on the surface and seek explanations for any given aspect of a system as a whole. Perhaps the reason that chemistry and physics seem so unromantic is the limited number of "things" to admire - there are only 108 elements we have discovered, vastly fewer elementary particles and only four fundamental forces, whereas evolution has given us an unimaginable variety of speciation and myriad ecosystems all over the planet.

Instead of picking a side in the reductionist/holist debate and dichotomizing aesthetic experience, I propose an integrative view to be superior. Observing the whole offers clues to the form and function of its parts, and understanding parts yields a greater appreciation for the whole. This approach can be done fractally, as we can take the parts to be wholes themselves, and examine instead parts of those parts. Once we "hit the bottom," so to speak, the whole is understood in terms of *relations* of its parts, and the relations between levels of observation.

The organization of simple parts that come together and interact to form wholes with properties that seem irreducible to those parts is called *emergence*. Note the word "seem" in the preceding sentence; emergence is all too often given as a handwaving explanation for unexplainable phenomena, and to assert that some other unexplainable force is responsible for organizing parts into wholes is functionally equivalent to believing in magic. The long standing debate on the question of the mind and its relation to the brain is oft resolved with such an explanation: the organization of neurons, relatively simple structures themselves (albeit not fully understood), connect together in a way that to some extent is already determined by genetics to produce a full experience of being. In this manner, consciousness is said to emerge from the brain, but all this truly indicates is that our understanding of the brain is still very naïve (or that consciousness is to some degree influenced by magic).

A more visible (indeed more tangible) example of emergent intelligence is found in the typical ant colony. Individual ants, like all insects, are not at all intelligent. Their 25,000 neuron brains (Tefl), compared to 10 billion in a human brain (Herrup & Williams), essentially lead individual ants in a genetic program, reacting to pheromones left by other ants. There are different pheromones for danger and for finding food, and the pheromones released by each ant are a probabilistic result of which pheromones they are sensing, how strong those pheromones are, in addition to other local factors including unforeseen obstacles. Ants coordinate navigation around seemingly uncrossable gaps by linking their bodies together to form bridges (Hofstadter). Knowledge of threat in one part of the colony quickly mobilizes the whole colony to defense, even if direct knowledge of the threat is outside the awareness of most of the ants. If a rock is blocking a pheromone path, ants will continue around the obstacle, initially moving either left or right with equal probability, dropping pheromones all the way. Those ants that pick the shorter path will find the original pheromone trail quicker, and will thus emit more pheromones, attracting more ants and rapidly reconstructing the trail as a result of feedback from the other ants that follow the pheromones (Dorigo).

This intelligent behavior has inspired an ant-based approach to a classic problem in graph theory, the travelling salesman problem. Graph theory concerns itself with mathematical structures called graphs, which consist of a set of *nodes* and a set of *edges* between the nodes. In the travelling salesman problem, or TSP, the edges are weighted and represent distances between cities, with the goal being to find the optimal (shortest) order to tour all the cities and return to the starting point. In the ant-based scheme, "ants" start at a node and "release" pheromones which is represented as a function on the edges of the graph. Ants walk around in random paths around the graph, dropping pheromones while attempting to make a full tour. Pheromone concentration is inversely proportional to tour length, and as the ants crawl around, shorter and shorter tours gain preference, and a near-optimal solution is found rather quickly (Dorigo).

TSP is a classic example of what is called an NP-hard problem, meaning that it is one of the hardest problems the answer to which can be verified in an amount of time that is polynomial in relation to its input. An n city TSP scenario has $\frac{n(n-1)}{2}$ edges and $\frac{(n-1)!}{2}$ possible solutions; this number explodes with the size of the input making brute-force methods for finding solutions impractical. Once a tour is found, however, it is a small matter to simply add up the distances between the cities in the tour and check it (it is actually linear, a set of n cities will take n computations, addition in this case, to check). Whether an algorithm exists to find the answer in polynomial time is still up for debate, and algorithms such as these form the crux of the $P = NP$ debate, an important open problem in computer science that also happens to be a Clay Mathematics Institute million-dollar millenium problem, which if true would mean that such problems have an algorithmic solution that gives the results in polynomial, instead of factorial, time.

Humans are surprisingly good at solving the TSP (MacGregor). For as many as 20 cities, human solutions tended to be within 1% of the optimal, with performance decreasing as the number of cities increased. While most humans don't spend their time looking for optimal solutions to TSP, we do similar searches on a daily basis - planning errands, for instance. You may want to do your shopping before you pick up the kids, you don't want to drive all the way across town then back to complete another errand, or there may be any number of preferences and constraints involved in planning your day. This natural inclination toward optimization led human participants to avoid crossing lines, a criteria that indicates a solution is not optimal, and to connect cities that lay adjacent to each other on the convex hull (if the cities were pins on a bulletin board, the convex hull is the interior of a rubber band stretched around them all), a criteria implicit in all optimization strategies. Interestingly enough, humans perform markedly better when presented with the spatial arrangement of the cities in question rather than a matrix of distances. Additionally, when humans were instructed to find the most "aesthetically pleasing" path, they performed worse than when trying to find the optimal path but still found solutions similar to optimal ones. It seems that we humans know what we like, and what we like is not working too hard for food. Evolutionarily speaking, being good at optimizing means less time travelling between food sources and more time actively looking for food, as well as minimized exposure to predatory threats. Indeed, when monkeys were tested with a variation of the TSP with several sources of food serving as the "cities" and the monkey's foraging scheme the "tour," they found near-optimal solutions as long as the number of "cities" wasn't too much bigger than 6, as the monkeys could not remember more.

Even brainless organisms exhibit a tendency for this kind of optimization. The lowly slime mold, which is not actually mold but rather a protist, is distinctive in many ways beyond its off-putting name. *Dictyostelium discoideum* in particular is notable; it starts its life cycle as a single cell hatched from a spore. When its own supply of food acquired during its vegetative state runs out, it emits a chemical that diffuses and attracts other nearby slime molds. As they bump into each other, their membranes fuse and they become a single entity, an amorphous slug that works as a colony to gather food. It travels along until it finds a food source, where it leaves a bit of slime connected by a vein to the rest of the mass that keeps searching for food. A Japanese research group laid oat flakes in a pattern similar to (in the geometric sense) major stops in the Tokyo subway system. They then let the mold do its thing, and by the time it had found all the flakes it bore a striking resemblance to the layout of the lines in Tokyo (Tero et. all). Allow me to reiterate: brainless slime mold was *almost* as effective at

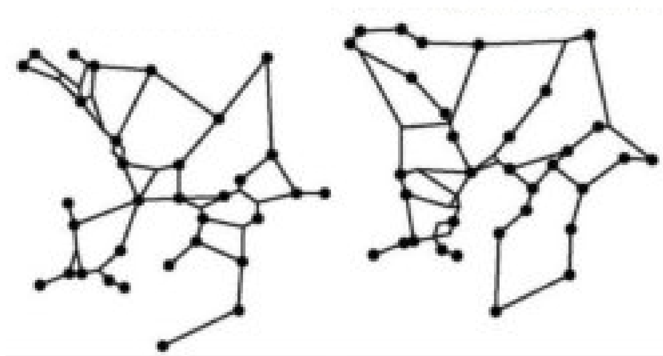

Figure 2: Left, the slime network. Right, Tokyo rail system map

designing an efficient transport network as a team of Japanese engineers. The slime mold certainly didn't plan anything though - through random mutation and subsequent evolution it found a scheme to solve a fundamental problem of life on earth.

The real hero in the realm of emergent life is DNA. The question of *how* DNA got here, the question of abiogenesis, is yet to be answered, but what really matters to us (for practical reasons, at least) is that whole beings can sprout up, with lives as rich as yours or mine, from packets of biological data. The human genome in its compressed, unexpressed state takes up around 650 megabytes (3 billion base pairs, four 2-bit pairs per byte) according to the Human Genome Project, and after a period of growth and learning emerge into the beautiful naked apes which have done so well for themselves on planet earth. DNA bridges the gap between emergent structure and emergent behavior, and examples of emergent structure come from all over. One that ironically gives us a warm-fuzzy feeling despite its naturally cold and prickly nature is the snowflake. Just like DNA (and therefore people), each snowflake is unique not for any particular reason other than the immense statistical unlikelihood of finding identical ones.

As a result of the weak forces between water molecules, itself due to the negative charge-heavy hydrogen attracting and the positive-dense oxygen, solid ice exhibits a hexagonal symmetry (Nelson). As tiny crystals of ice form they start to fall to the ground, pulled along by the wind and gathering mass in spurts and stops as it travels through temperature fluctuations in the upper atmosphere, accumulating on its limbs more ice, getting larger and more complicated. By the time it reaches us on the ground it has taken the shape of the six-fold flake we know so well. Popular conception of what snowflakes look like is due mostly through the work of one very dedicated snowflake enthusiast named Wilson

Bentley. In order to study snowflakes without them melting as examined them under a microscope, he perfected a technique catching them on black velvet and photographing them with a bellows camera attached to a microscope. The resulting thousands of pictures provided enough detail that "hardly anybody bothered to photograph snowflakes for almost 100 years," (BBC) and it was he who originally proposed that no two are alike. Bentley died at age 66 after catching pneumonia, which to nobody's surprise he contracted after a long hike in the middle of the winter to photograph snowflakes. His short biography is included as I had never heard of him before, and his contributions to human knowledge, regardless how niche, deserved acknowledgement.

Another snowflake of interest to mathematicians exhibits a different kind of similarity than rotational symmetry. Indeed it is unrealizable in our world - the Koch snowflake, which is defined recursively as follows: beginning with an equilateral triangle, take the middle third of each side and add two more line segments to create an equilateral triangle that points out. Delete the original middle segment and repeat the process with the three newly created triangles (Burns). The Koch snowflake is the limit of this process, unrepresentable on even the largest sheet of paper with the finest pencil. If you attempt to draw just a portion - the tip of a tip of a tip of a triange - the process still goes to infinity, and the result is identical no matter how small of an area you examine. The Koch snowflake is an example of a *fractal* - a concept first coined by Benoît Mandelbrot in 1975. Formally, a fractal is a set whose *fractal dimension* exceeds its *topological dimension*. The latter is what is usually meant when we say "dimension" - how many coordinates are required to define a single point.

Side note: the consideration of dimension above applies specifically to euclidean space. Coordinates are given in terms of linear combinations of vectors. One could easily define *four* coordinate vectors to any point in 3-space, but at least one vector is redundant and can be removed. The *minimal* number of these is the dimension n, and any set of n vectors that can be used in a linear combination to describe any point in the space are called a *basis*. Most people prefer working with *orthogonal bases*, meaning the vectors are all at right angles to each other, because they have some very nifty properties and more practically encapsulate the notion of dimension with which we are all familiar. This notion is generalized to other spaces with the notion of open covers - enclosing a space in the union of some open sets (sometimes said a *collection of sets*), the basic elements of a topology. The dimension of a space is n if and only if for any open cover, a subset of the collection (a "refinement") can be found such that no point is in more than $n+1$ sets in the refined cover.

So in the example of the Koch snowflake, its dimension is one - both because it is defined as a set of line segments, euclidean notions of dimension one, and also because an element only has to be in at most two open intervals that cover the snowflake. To say that the fractal dimension exceeds this is like saying that

Figure 3: Some fractals. Left to right: Satellite image of a watershed, Romanesco broccoli, and a bush generated by an L-system

the infinitely many line segments are too tightly wound up that it begins to look surface-like at the edges, but with too many “holes” to really be a surface - in this sense, its dimension is *between* one and two (Mandelbrot). This is the trait that endows fractals with their infinite-yet-finite nature. Not all fractals have to be defined recursively or possess self-similarity, but those seem to be the easiest to define. Recursion enables a simple set of rules to break the whole down into smaller bits, and in the case of the Koch snowflake the property of self similarity comes from the fact that the same rule is being applied at every level. Famous also for its self-similarity is the Mandelbrot set, which is also defined in a recursive way by iterating a complex polynomial and checking if its limit is bounded. It is not exactly self-similar but different levels and the patterns that arise within could be considered variations on a theme.

So recursion allows us a way to coil up a line so that it has some semblance of area. This concept plays a huge role in the design of transport networks, both human-made and natural. Prominent is the fractal arrangement of plant life - branches grow at angles from the trunk, which themselves branch from other branches, which grow leaves at angles from the branch, which themselves have branching veins within to move water and sugar around the leaf. Cauliflower branches in groups of about 13 branches, and romanesco broccoli exhibits an even more striking fractal arrangement, so much that it looks somewhat alien. Ferns in particular bear a remarkable self-similarity, with each “leaf” in the fern a composite of many smaller leaflets, arranged in the same fashion as the whole. What this arrangement means, especially in the leaf, is that the compromise of one pathway of the tree system does not spell demise for the whole (Lindemayer). Punching a hole in the middle vein of a leaf will not compromise the whole leaf’s ability to dissipate water, as water will still be transported throughout the smaller veins in the leaf (Katifori et al.). Furthermore this approach has the advantage of being easily encoded, especially relative to the final, emergent structure.

In his work on algae growth, Hungarian botanist Aristid Lindemayer came up with a formal system to model the growth of algae. Using a simple two-stage interpretation of algae growth and a set of rules for getting from one stage to another, he described how multicellular forms emerge from these simple rules. Called an L-system after Lindemayer, it can be used to draw several fractals that are defined recursively in other ways. The Koch snowflake, for instance starts with the triangle and recursively adds more triangles, changing the boundary by changing the interior. With an L-system definition, however, each iteration draws a different boundary, and operates on the boundary from the previous level of recursion, using orientation of the line segment and a specified angles to generate the next level. This method can also be used to draw convincing-looking plants, if an appropriate set of rules governing angles and replacement is selected (Lindemayer). L-systems can also be used to model terrain. As Mandelbrot said, "mountains are not cones," and computer generation of believeable mountains and terrain incorporates L-system rules with a bit of randomness we have come to expect from nature to generate these patterns. This technique was first exhibited in 1980 by a man named Loren Carpenter, who would eventually go on to co-found Pixar, and was used in the *Star Wars* films to generate the surface of the Death Star.

5 Maps and Territories

Scientists make regular use of models and simulations in the course of their work. A model in science, just like a model airplane or action figure, is a representation that captures the properties for which the whole is important. A model airplane has a fuselage and wings (if it did not, I would be hesitant to call it a "model airplane") and some exceedingly detailed models may include little cockpit controls and detailed seats on the inside and maybe even a little pop-open cargo bay. If said model were absurdly detailed, it may have real (tiny) working engines and the controls might actually be controls, and connect to all the functional pieces and control them. The limits of our engineering to make tiny bolts and screws to fit in our model might soon be reached the more detailed and accurate we try to make the model, however. If we wanted to make the best model airplane possible, we would have to keep enlarging and adding more detail until the next thing we know we're left with is a life-size, fully functional, "model" airplane.

So what we are left with is scarcely a "model" anymore; if i wanted to collect such model airplanes they wouldn't fit on my bookshelf as they did when I was a child, I would need a hangar and more aluminum than I knew how to obtain at the age of eleven. I wouldn't have minded of course, but given the infeasibility of actually collecting planes, when we're talking models the question of "how good is good enough" invariably comes up. To some simply wings, a fuselage, and tail assembly is sufficient, and if it manages to look like an airplane that's even better. Some may wish for greater detail in certain areas they perceive to be important, but no "model" airplane can ever be 100% accurate (and nobody expects a model to be).

What, then, is important in our airplane model? Since "important" is a subjective judgement, it of course depends on what the model will be used for. If you need to show your young nephew what an airplane is, all it really needs to be is a tube with wings and he will likely get the picture, and in his mind it will resemble a boeing 747 (because it's pretty much just a tube with wings). If he wants to know how the plane turns in the sky, it will need to have wing flaps. If he wants to know where the paratroopers jump out, then it will need a moving cargo bay or some such. Hope he doesn't want to know how to build one from the ground up, then you may want to eschew a model altogether and just give him a set of schematics, with diagrams for every functional part and a layout of where each nut and bolt goes!

So we could say that a model is "good enough" if it can convey those aspects of the original that we care about. Surely a model plane with thousands of tiny nuts and bolts would be past the level of detail which most (reasonable, non-aerospace engineer) people care about, but one lacking any sort of tail would not convey important details about stabilization and dynamics of controlling the direction of the plane. Since controlling the direction of the plane is an aspect of the fundamental *purpose* of the plane, namely flying, whereas descriptions of each nut and bolt would be data describing, essentially, that some parts of the airplane should be held together, which is nearly self-evident. Here it is important to note that while a *model* can indeed be a tangible object and appear similar to what it represents, like a die-cast hot dog with wings, but a set of complete, bolt-by-bolt schematics *also* counts as a model. Instead of a scale representation, where the "encoding" from original to model is essentially a scalar transformation, the schematic representation is more conceptual, encoding 3-dimensional information in a series of 2-dimensional pictures and text which contain enough information for the original, tangible form to be reconstructed. Deciphering such a schematic encoding requires a knowledge of not only what a 2-dimensional projection of a bolt looks like, but also the context of knowing how to decode information about distance and scale from the schematic (the skill of reading a technical document) as well as the proper linguistic background required to decode the accompanying notes and instructions.

The vocabulary of functions as "maps" is reminiscent of physical maps - the folding pieces of paper that are a picture of the world, if all the water in the world was uniformly blue and the land uniformly green. By possessing those qualities, one can deduce, from seeing a multitude of other maps with similar attributes, that the folded piece of paper is a map. It may even say something like "San Diego" on it, which would be a dead giveaway since you cannot obviously hold San Diego in your hand, hence must be a representation of San Diego. The paper may come with its own decoding system, called a key (which must itself be decoded using language faculties to establish a correspondence between symbols, like a dotted line, and symbolized, like a road) and some indicator of scale. The correspondence between the two is strong - locations of places of interest (important places) will be on the map and be arranged in a manner that corresponds to dots on the map. A certain distance travelled will correspond to a certain distance on the map, no matter where on the map you may be (this is of course not true for maps of large scale, as flat projections of curved spaces always have an inherent distortion of distance). There must always be some sacrifice of detail, of course. The small but nonzero resolution of printers, and hence the resolution of the printed map, corresponds to a scale resolution in the reality it represents and everything beyond that is lost, but if we try and expand our map and reduce our scale to include what we originally missed we will never be able to include everything. Furthermore the map just gets bigger and bigger and more cumbersome and even harder to hold in your hand, and the next thing you know you're hauling a piece of paper the size of San Diego in the back of a semi truck, and then the convenience of having a map is lost. In the words of Mein Herr from Lewis Carrol's Sylvie and Bruno, when Sylvie remarks that six inches to a mile is the largest useful scale:

> Six inches! We very quickly got six *yards* to the mile. Then we tried a *hundred* yards to the mile. Then came the grandest idea of all! We actually made a map of the country, on the scale of a *mile to the mile*! It has never been spread out, yet. The farmers objected; they said it would cover the whole country and shut out the sunlight! So now we use the country itself as the map, and I assure you it does nearly as well.

While a map or model may never be able to contain all the information about the original, it is important to note that *this is not a bad thing*, and in fact is the reason models are so handy - they take an infinite amount of information and extract that information which is helpful to know. So instead of carrying a map the size of San Diego, you have a map that fits in your pocket that may not detail where every water fountain in downtown is, but it gets you where you're going and hey- *it fits in your pocket.* Likewise with your brain - it's prone to error and certainly cannot know everything, but it carries a representation of reality that fits inside your skull! The important thing is to recognize the model for what it is - a tool for understanding a greater, complicated whole but not as a stand in for the whole itself. In the immortal words of Alfred Korzybski, "the map is not the territory."

By rewording our definition of "model" above in the framework of encoding and decoding, we can say a model is effective if it is encoded such that important information about the original can be extracted using relevant knowledge, heuristics, and other decoding mechanisms expected to be possessed by those who would be using the model. For instance, a general understanding of large machines and that they are typically made of metal, held together by welds or nuts and bolts, and rely on some sort of engine to drive them is a categorization that allows the observer of the model plane, by simply knowing that it is a machine, can infer that much about the plane and thus the model plane not need be made of metal, held together by nuts and bolts, or have a working engine to be an effective model. That the plane is a "machine" is itself an encoding; that is to say that the word "machine" is an encoding of ideas, concepts, and experiences regarding machines into a written system, a written system that is itself an encoding of a spoken language into arbitrary symbols. Hence linguistic competence is another important aspect of the relevant knowledge that goes into being able to decode the meaning in the model.

A model need not represent a physical object, either. Scientists use models to explain complex or abstract processes in a manner that is accessible to others. In this sense, using a model is a great advantage, because the *important* aspects of the theory which the model is an attempt to convey are framed in a manner such that it can be decoded by others' relevant decoding systems. Ask anyone to draw an atom, and they will likely draw a dot (or perhaps dots) to represent the nucleus and some concentric circles with dots on them to represent orbitals and electrons. That is not what atoms "look" like though, since electrons are massless, orbitals are actually probability clouds, and the distance from the nucleus to where an electron is likely to be is so relatively huge that if the nucleus was the size of a penny, the atom itself would not fit inside a football field. The real killer is thinking of the atom as "massive" without giving any thought to the fact, originally shown by Einstein, that mass is simply condensed energy. Though the model that many people have is lacking in several areas, it is sufficient to explain a variety of nuclear and electromagnetic phenomena that we can observe, and so for most people the model of the atom is "good enough," and indeed its explanatory reach is much greater than earlier models of the atom, J.J. Thomson's "plum pudding" model for instance, in which the atom is a field of positive charge, like a pudding, with little negatively charged plums interspersed, thereby totally missing the concept of the nucleus or energy levels. In this way we can say that the incomplete yet popular conception of the atom is better than the conception held at the turn of the century.

Not only does a model not have to stand for something physical, but the models themselves need not even be tangible. That is to say, a model can exist inside the mind, completely inaccessible to anyone without *another* encoding step required to externalize it. I say "completely inaccessible" in the sense that a. The encoding-decoding system for thoughts is inextricably linked to individual experience, which influences neural pattern growth and hence the linguistic, visual, and ontological neural structures and b. Any attempt to extract information directly from the neural substrate would be simultaneously monumental in its difficulty (in the computer-scientific notion of the space needed to store said neural information and the processing power needed to keep up the encoding in real time) and incredibly invasive in any conceivable means. Since this is how *all* conceptions of *everything* we perceive and integrate into our worldview are represented to our minds, our theory of how reality is organized and how it operates is remarkably individual.

Knowing this about your brain is also a key to overcoming a mental hurdle set in place about evolution. Animal (and thereby human) intelligence evolved to increase chances of survival, and as such the human mind is constantly looking for survival problems to solve. The level of consciousness that is linguistically aware of the neural activity taking part in the other areas of experience, such as the motivational or sense-perceptive areas, the "inner monologue" or "ego" is constantly performing the process of accessing and modifying those structures involved with "knowing" things about the surrounding world, optimizing our actions for survival and procreation. By being aware of this process, and realizing one's own thoughts are simply aspects of the various models regarding reality coming to awareness, one can overcome a sense of certainty in the workings of the world around, and with training (like the meditative practices found in Buddhism) can quiet these now-unnecessary survival mechanisms and allow one to exist independent of judgement and categorization.

6 Evolution

> "Why do gorillas have wide nostrils?"

As a resident of the great state of Tennessee, I am proud of my state and its long history of trying to keep dangerous scientific theories out of the hands of impressionable youngsters. In setting a tradition that would endure to the present day, the 1925 Scopes trial held in Dayton ensured that never again would anyone stereotype the average Tennessean as scientifically literate. Keeping with tradition, current governor Bill Haslam recently signed into law a bill that would allow science teachers to teach things other than science and not worry about getting fired - namely, schools of thought (I hesitate to say the word "theory") antithetical to the current prevailing trends in the academic community, especially those regarding climate change and evolution.

Actually, it is a tradition that should have made Tennessee a laughingstock of the country, and it would have if the antiscientific sentiment in this country weren't so unsettlingly widespread. In my experience, these beliefs have several causal factors. First and foremost is a literal interpretation of holy texts in the Judeo-Christian traditions, and the political theatre in which these religions have carved out a significant part, both of which are (thankfully) outside the scope of this paper. What certainly is not out of the scope of this paper is the other side of the coin - that is, incorrect biases and notions about what the *theory of evolution* means, namely misunderstandings about what *theories* are and what *evolution* is.

If there is one epistemological notion that i've tried to drive home, it is the notion that absolute, unquestionable, unfalsifiable knowledge does not exist. Like our example earlier, if two people have contradictory opinions or observations, then at least one of them is wrong, but if they agree then there is no knowledge gained - they just agree. A fundamental part of the scientific method is coming up with reasons to explain observable phenomena, called hypotheses, and then trying very, very hard to disprove them. For if the data says one thing, and the hypothesis predicts another, then either the data is wrong or the hypothesis is. In undergraduate lab classes, it is most usually the former, and in the world of "real science" it could be either. Earlier this year data obtained about faster-than-light neutrinos stood at odds with the universal speed limit on matter detailed in Einstein's theory of relativity. The "bad data" that led researchers to this conclusion was able to be reproduced, leading many people of lesser faith in the veracity of relativity to expect a monumental new discovery, when in fact the FTL neutrinos were more a result of glitchy measurement tools than anything else.

Part of the reason proving the "theory of relativity" wrong would be such a big deal is that in order to be a theory, a hypothesis must have so much explanatory power and be able to withstand a great deal of rigorous scientific testing. This stands in contrast to popular usage of "theory," as in "my theory is that Bob dumped Susie because he couldn't stand her sister." That statement is neither falsifiable nor does it have explanatory power, and it probably wasn't subjected to any rigorous testing either. So when people can have "theories" about any phenomenon they care to name, then what should be an awe-inspiring statement about the world gets interpreted in the minds of the masses as the idle musings of some scientific bigwig. This is a symptom of greater scientific illiteracy in our nation and, clearly, is not helping things along.

Possibly even more detrimental to people not accepting evolution is a fundamental misunderstanding of what evolution is. The infamous argument likening evolution to a whirlwind travelling through a scrapyard and leaving a 747 in its wake is so laughably out of touch that it catches people off guard. That some would accept "evolution" only if that evolution was guided by some architectural force in the universe need to take a close look at the platypus and all its absurdity, or the human retina that's installed upside down (especially given that there are other animals who get it right). And if you believe at all that "evolution" is anything but a statistical result of playing the game of life on earth, if you believe evolution to have a purpose of any kind, then you are doing yourself a disservice and missing the point.

Evolution is the gradual change of species over time as a result of certain traits possessed by individuals contributing to greater reproductive success. It is the cruel reality that life is tough, and some individuals just don't make it. Whether it's a dramatically low chlorophyll count, smaller thighs, or a reproductive tendency to lay eggs at a rate smaller than what would ensure reproductive success, if an individual is not *fit* enough to survive in its ecosystem until it can reproduce then it doesn't get to. What ends up happening is traits that promote adaptation are passed along to future generations, thus increasing the proportion of the population with said trait, propagating it generation after generation until said beneficial trait is prevalent throughout the whole of the population. I have been dancing around it, but *generation after generation* and *future generations* should have clued in the reader that this change is gradual. Gradual in the in the sense that it takes a large number of generations for a gene to spread through a population, and in the sense that mutation, the driving force behind evolution, is a random process, and not all mutations are beneficial.

Although mutation is random, and evolution has no direction, for many different biological processes we have formulated many corresponding mathematical models, and often a result regarding optimization in the model corresponds to an actual, evolved trait. Spiral packing, for instance, is the best way to pack round things like seeds or rods and cones from a single budding agent (Ridley), and as a result this form (in the Platonic sense) often finds itself manifested in the natural world. Did the *first* seeds in the proto-sunflower arrange themselves in a spiral? Probably not, since of all the ways to pack seeds there are many that will do a "good enough" job. It only takes one exceptionally harsh season though to wipe out or otherwise deny the opportunity to reproduce to all the proto-sunflowers except those most efficient at producing seeds.

While evolution is directionless and purposeless, more often than not a mathematically optimal "solution" for some given trait exists. Thanks to the scale of time that evolution operates on, a species can come to approximate that solution quite well simply because the optimal solution is by definition the best way to ensure reproductive success. The word "approximation" is reminiscent of computer algorithms to find solutions to classic problems in algebra and analysis - namely, a "guess-and-check then make a better guess" system of obtaining a machine approximation to a real-valued solution. Compare the process of evolution to a polynomial root-finding algorithm that employs the bisection method:

Evolution	**Bisection Method**
Organism is born	Solution exists in interval (a, b)
Phenotype manifests	Midpoint $c = \frac{(a+b)}{2}$ is calculated
Fitness is tested	Test for solution in either (a, c) or (c, b)
Organism yields offspring	Interval with solution becomes new (a', b')

The testing stage is where the interval containing the solution is refined, and the testing stage where an organism's fitness is decided. It is where the environment influences genetic progression, and in this sense evolution can be viewed as a dialectic process. Similar to the concept of a dialectic argument in classical thought, where two parties with two differing points of view exchange ideas and reasoned arguments in hopes of reaching a conclusion that contains some level of truth (contrasted with *rhetoric*, getting people to agree with your point and you're the only one who gets to talk, and *debate*, where neither you or your opponent cares about the truth, you only care about winning the argument), a dialectic process is one in which *part* and *whole* (organism and environment in this case) influence each other such that the causal borders between the two are blurred. The evolution of the organism and the evolution of the ecosystem are deeply linked together; if evolution builds a better fox, then it will also build a better rabbit, or else there will be no more rabbits. Ecosystems evolve with every species' evolution dependent on the evolution of every other species, whether that dependence is superficial or profoundly deep.

The question of how humans evolved, then, is tied deeply to the ecosystems where early humans found themselves and is a direct result of the evolution of ancestor primates (and of mammals before that, and of animals before that...). Climbing down from the trees and into the grasslands meant that those hominids that could stand upright could see farther, spotting predators and prey from far across open grasslands. Evolving simultaneously were our dextrous hands - not using them for climbing and walking meant we could use the to hold, carry, and manipulate things, allowing us to use tools and increasing our brain size as a feedback result. Larger brains means larger heads, and even though females developed wider hips, a great deal of human cognitive development takes place after birth, unlike in many animals. Eventually our hairlessness an bipedality gave us the upper hand in endurance running, and our big brains allowed us to predict and track our prey, and combined with our ability to harness fire and cook food opened up a whole new world of calories to be had (Larsen).

Most importantly to our survival, perhaps, is the fact that humans are social animals. All mammals are "social" in that some level of bonding between mothers and offspring takes place (indeed this is a hallmark of mammalian child-rearing), but all hominidae, great apes including chimpanzees, gorillas, and of course, humans, have complex social structures that determine how a species manages mating pairs, child rearing, and gathering food. The evolutionary benefit of being a social animal is clear: working together, a group can accomplish more than an individual when it comes to hunting and self-defense, and cohesiveness of the group is important to its success. So mammal, and in turn human, cultures that exhibit a more cohesive structure would do better than competing groups, so traits which contribute to a group's cohesiveness will be selected for.

A distinguishing feature of human evolution that contributed to a larger social structure is our remarkably large brain. While the sheer size of the brain would allow for a more effective long-term memory, resulting in stronger social bonds between individuals, certain areas of the brain grew more proportionally than others - namely, the temporal areas of the brain, used in language production and understanding as well as long-term memory, and the prefrontal cortex, which governs social behavior and is associated with higher-level decision making (Schoenemann). This is an example of a trait evolving dialectically - a larger capacity for complex social behavior lends itself to tighter social structures, which in turn lead to more effective societies with regards to hunting and protection, meaning individuals within the group who navigate the social structure better have a better chance of producing offspring as a result of sexual selection. This would, in turn, put an evolutionary pressure on still larger brains, which is calorically costly (some 20-25% of the calories consumed by a human are used in keeping the brain working, Shulman et al.) but will yield a payoff in the long run as the group becomes stronger.

When humans developed the capacity for abstract, symbolic thought and communication, it meant more than just better organization of our social structures. It also gave evolution new organizational levels on which to operate. In Darwin's conception of evolution, evolution as it applied to biological systems, genotypic differences resulted in corresponding phenotypic differences, which were then selected for by forces in the ecosystem. Evolution doesn't necessarily have to operate on the genetic medium, though. If elements (elements being the fundamental atomic units that combine to form greater structures, as atoms are with massive objects) of a system can be transmitted with near-faithful accuracy, allowed to change in successive iterations of transmission, and exposed to some sort of selection pressure, then that system will be subject to evolutionary change.

Such is the case with memes - units of cultural as opposed to biological inheritance. Coined by Richard Dawkins, the evolutionary biologist more known for his attacks on God and organized religion, memes can be songs, sayings, ideas, or tropes, and they disseminate not through sexual reproduction but through communication between people. And through that communication, by jumping from one mind to another, the meme changes slightly, or mutates, as different individuals give it different consideration. They link up and produce offspring, they compete for resources, consideration in the minds of men. The most successful memes are the ones that get adopted and integrated into the greatest number of minds. Fads and larger cultural trends are memes that survive because the current cultural state of a society is indicative of minds ready to adopt those beliefs.

Dawkins introduces the concept of "meme" in his 1976 book *The Selfish Gene*, and as early as then he makes the case for a Darwinistic view of God - that is to say, nothing whatsoever. The snippet of information "Jesus is the true Son of God" is a very powerful meme - it sits at the core of religious belief for a third of the world's population. Like a virus, this concept incubated in the minds of the members of the early church and spread as it came in contact with more suitable hosts, lying dormant in the written word of the bible, and exploding as it became

Figure 4: Some memes, left to right: unexplainable love of bacon, first four notes of Beethoven's Fifth Symphony, aptly named Ugg boots, a meme as memetically interpreted by the internet.

the de facto religious belief of most of Europe. The rest, they say, is history. In *The Selfish Gene* Dawkins argues that "God" is a powerful memetic concept. Indeed it is: a belief in God answers a lot of unanswerable existential questions, like "why are we here" and "what comes after death," making it a lucrative concept to believe in hence giving it great memetic potency. Eternal damnation and blind faith are other classic components of Christian belief systems, also memes that contribute to Christianity's staying power. Faith is an especially potent meme; nothing contributes more to a meme's fitness than immunity to critical consideration.

The problem with this analysis is that Dawkins is memeticicizing components of a religious experience while ignoring the fact that spiritual experiences are characterized by their ineffability. By spiritual experience I mean an anomalous, ecstatic experience where "worldly" concerns and states of being yield to a feeling of connection with the "divine." They are ineffable because engendering such a state is no easy task, as such the potential for shared experience is rather low compared to day-to-day experiences like eating or sleeping. Furthermore, because of their fleeting nature, the full scale of the experience is near impossible to comprehend let alone convey. A meme, on the other hand, must be encoded into some symbol-system in order to propagate and spread to others. The purpose of religion is to systematize a way to interact with the divine, and it is religion that is full of those powerful memes. True spiritual experiences are outside of the realm of symbol-systems, and any attempt to describe them is incomplete at best, woefully misleading at worst. Dawkins makes the mistake of disproving the existence of God by claiming that a memetic conception of God is an entirely human construct, which is absolutely correct but does nothing to support his end goal of disproving the existence of God.

Taking the memetic concept to the next level is a project called DarwinTunes (MacCallum et al.). Calling itself "survival of the funkiest," it is an internet-based project that crowdsources (a lovely neologism meaning "make the users do all the grunt work") the process of selection with the hopes of creating listenable tracks from noise. Starting with what essentially sounds like static, each generation consists of a handful of loops which "reproduce" and undergo "mutations," resulting in a daughter track that combines the two with random changes in the sonic signature. Users rate the tracks, and the rating serves as a basis for a fitness function: the more ratings a track has, the more it gets to reproduce and mingle with the others. Each track is algorithmically determined; it is not encoded like a MIDI file where each note corresponds to a different byte in the source file, but instead is produced from a code. This further helps the analogy of evolution by way of phenotypic representation from genotypic information. After just a few hundred generations, the loops begin to take attributes of human music - rhythm, melody, and even a coherent key are eventually manifested in the fittest (funkiest) loops.

Besides demonstrating memetic theory in a very tangible way, DarwinTunes is one aspect of a growing trend in selection that began when we first harnessed nature to do work for us. When we stopped having to spend all our time hunting and gathering, and settled down to specialize and develop societies and economies, the time it took for a gene to propagate across a population paled against the speed of memetic transfer. Thus began a separate track of evolution - cultural evolution, that changed far faster than the biological substrate supporting it, the human population. When culture progressed to the point that it could support the rapidly improving technologies of the industrial revolution, this new evolutionary/memetic track of technological evolution could begin that far outpaced the comparatively sluggish cultural track supporting *it*, to the point of technological innovation driving cultural change. What is happening is a telescoping effect of evolution (Linklater), with the selection and propagation times getting progressively shorter and shorter and evolution happening correspondingly faster. The internet as we know it enables the quickest, most widespread dissemination of information, and the resulting evolution of what is essentially expression and knowledge is happening at a breakneck pace, influencing all the layers below it - technological, cultural, and soon, biological.

And if you haven't guessed it yet, the answer to the joke at the beginning of the section is big fingers.

7 Brains

A distinguishing feature of animals, save a few cases of invertebrates, is the brain. It generates the conscious experience of fish, reptiles and mammals, as well as makes all the thinking and talking which we humans pride ourselves on possible. While behavior varies wildly across all animals, the basic functions of the brain orchestrate the various drives for survival, and these basic functions are reflected in analogous structures that appear in all vertebrates. Decision-making, eating, reproduction and other motivation based activities are focused in the forebrain, which develop into structures in humans that include the frontal lobes, areas crucial to abstract thinking and planning (Squire). Movement and autonomic cycles like sleep are controlled by the midbrain. It is also where dopamine, a neurotransmitter found in all vertebrates that governs arousal and reward circuitry, is produced. In the hindbrain are the definitively unconscious areas that control breathing, heartbeat, bladder control, and is important in movement, as well as vision.

Unique to mammals, however, is the neocortex, which is markedly larger in humans. Distinctive in having six layers, it is a sheet of grey matter that is folded upon itself, which yields the practical benefit of having a larger practical surface area (and filling space for that matter - the human brain has a calculated fractal dimension of about 2.8, Liu et al.). The neocortex is divided into two hemispheres, with two corresponding sets of four lobes, which are separate both in function and in the larger grooves of the brain. The parietal lobe is responsible for keeping track of the body in space - it takes sense data from muscles all over the body and creates a coherent self-image. Data from the vestibular system allows the brain to determine the orientation of the body in space with respect to gravity using a set of three semicircular canals filled with fluid, oriented at nearly right angles to the others. The magnitude of fluid moving through these canals enables behavior essentially analogous to an orthogonal basis that spans 3-space - every human carries such a basis (two, actually) around in their head!

By virtue of its proximity to the occipital lobe, the parietal lobe contains a number of areas responsible for mapping coordinates to objects in the visual field, and vice-versa it provides a sense of how that body image fits into the world around it. The process of assembling this image from a variety of sensory inputs takes place in regions called association areas, where stimuli are compared, identified, and given some sort of meaning of relation to the body. Skilled practitioners of Transcendental Meditation can enter an altered state of consciousness in which activity in certain areas defining the body's outline, namely the orientation association area in the left hemisphere, is diminished (Begley). The other main function of the parietal lobe is initiating muscle movement, which is modulated and fine-tuned by the cerebellum, which is tucked underneath the hemispheres behind the spinal cord.

Located between the cerebellum and the parietal lobe is the occipital lobe. Named for the occipital bone that rests over it, it is notable for being the lobe that is responsible for vision. The process of vision starts, like many of our other senses, in the external world (of course, if we take the word "sense" as any information that exists outside of our brain as opposed to our bodies then all senses are of this variety). Photons travelling into our eye come in all varieties of wavelengths, but the tendency for light in the 390-750 nanometer wavelength range to be "visible light" is not an intrinsic property of that portion of the spectrum but a reflection of the wavelengths that activate the photoreceptive cells in the eye. The eye contains two kinds of cells - rods, which are responsible for gross distribution of light intensity and are more dense near the periphery of the retina, and cones, which detect color and fine movement, hence their more packed placement near the center of the visual field. Cones come in three flavors - red, blue, and green, and all of the millions of perceptible colors come from adding these primary ones in different amounts.

The resulting sensation of color is a part of a larger philosophical debate over the nature of *qualia*. Like so many other sensations, the way color is represented in consciousness is unexplainable with our current understanding of our neurology. Textures of things we touch are felt as smooth or rough, smells are sweet or foul, and the apple on the table has a definite redness. Any such stimulus that is represented as some sort of feeling in the brain is a quale. The difference between touching and feeling or looking and seeing lies in the brain, where sensation is, linguistically and psychologically, made sense of. The problem might be understood when we can establish to what level our neurology corresponds for certain sensations. We can speak and associate and agree on labels for qualia that match labels for stimulus, but the problem of whether two different humans perceive the same redness of an apple is an untestable one because of the impossibility of any individual experiencing any reality outside of the one generated by their particular brain. The question of whether the quale is *inherent* in the object being perceived is a different question, one which i think is slightly easier to tackle (though again, as empirically impossible to verify). I discuss it here because the visual system offers several clues.

First is the fact that the sun, as many believe it to be, is not yellow. It is actually white (Cain), with the atmosphere reflecting the blue wavelengths making it appear yellow by the time its photons reach us. The sun is known across time and culture as being the single largest source of electromagnetic radiation on planet earth, and things that emit large quantities of radiation in the visible portion are translated in our experience as being very bright. In all our evolutionary history, our color vision evolved from sampling wavelengths from the sun, and *white*, the feeling of brightness, comes from the color spectrum we observe from the sun by stimulating all three kinds of cones. If an alien species evolved on a planet whose sun was what we would call red, then its vision would have, in all its evolutionary history, been sampling a "redder" sun, but there is no reason to believe that its experience of "bright" would be fundamentally more different than ours. They would certainly see our sun as having some brightness by virtue of being a stellar furnace and the accompanied radiation, but they may not see it as white.

The yellow appearance of the sun and the related blueness of the sky reflect a property of color of academic interest since Newton's time and a fundamental part of the curriculum for kindergarten, the color wheel. When we stare at an object of a certain color and then turn to face a white surface, and afterimage persists in the complementary color. For red, a cyan image remains, and blue leaves a yellow afterimage. Green appears as magenta, however - a color which does not have a corresponding wavelength on the visible light spectrum. That is to say, there is nothing that is *magenta*, only things that emit primarily blue and red colors (Elliott). Our brain invented it as an appropriate midway between red and blue, and our perception of it as a color enables us to take the segment of the visible light spectrum and turn it into a wheel, filling in smoothly a gradient between the primary colors of our vision. While magenta is the motivation for this discussion, in reality any color that isn't red, blue, or green is a color we have invented in our minds to fill in the gap. If none of these colors have any external source when they are represented in our minds, then there is no real reason to believe that red, blue and green exist outside of ourselves either. Again, the essential question is unresolvable, at least at the current level of our understanding, but understanding how and why our brain uses color are tied deeply with understanding how vision works in the brain and how we use visual information.

Returning to our whirlwind overview of visual processing, it is the occipital lobe that is mainly responsible for synthesising the cascade of individual firings of rods and cones in the retina into a coherent picture, identifying object and ground and differentiating between objects rather than incorporating that information into model of reality generated in the parietal lobe. The primary visual area, called V1, contains clusters of nerves that correspond nearly one-to-one with photoreceptors in this visual field and are tuned to certain visual cues, like colors, orientation, and intensity of the stimulus. Further processing takes place in the secondary visual cortex, V2, which identifies basic shapes from the constituent parts, and then signals get sent to the temporal lobes for recognition and the parietal lobes for integration into the worldview. Computationally speaking, recognizing an object by doing a simple search of your inventory of "things" and comparing it to what you have in your visual field is far less computationally effective than establishing a class of qualities that restrict that search. When you see a glass of water, you notice the straight edges of the glass, some contrast where the water line is, and colors that are the same as the ground. A basic shape of the whole takes form by the time the signals leave the occipital lobe.

The last of the basic senses, audition, is managed by the temporal lobes. In addition to the auditory adaptations we share with other animals, like listening for predators, the human auditory system is wired to recognize speech. The specific areas involved with language were recognized in the mid-1800s by studying the brains of people who had langage defects. Carl Wernicke studied patients with aphasia, an inability to comprehend speech. Lesions in the rear part of the temporal lobe led him to conclude that language comprehension is centralized there. Through a similar approach, Paul Broca examined patients who suffered an inability to produce speech after suffering damage to a particular area near the frontal lobe. In addition to production, damage to Broca's area also can result in damaged syntactical ability. Notable also is the role of the temporal lobe in forming memories. It contains the hippocampus, which governs memory formation, and the location in the linguistic center of the brain which contain the structures involved in deciphering the semantics of language may contribute to that. The actual ways in which visual and linguistic information are involved in the formation or structure of memory and knowledge have not been definitively made.

A notable feature of Broca's and Wernicke's areas is their localization to a dominant hemisphere, usually (97% of the time, Reeves) the left. This is the most notable example of hemispheric specialization, and every other example seems to be related to the faculties of processing language. Understanding and producing language is largely a task of encoding and decoding symbols along certain rules. This symbolic manipulation is abstract, with choices of vocabulary and grammar being mostly arbitrary. Thus the left hemisphere of the brain *tends* to specialize in abstract, formal, and categorical thought. The right hemisphere shows a slight tendency for specialization in gross, holistic processing, such as spatial reasoning and visual processing. A couple of notes on these observations: firstly, beyond the functions of the major language areas, neither left nor right truly *specialize* in logical or creative thought, as the dichotomy is often popularly understood. It would truly be a sight to see a "right-brained" person who had not undergone surgery. Secondly, even in the result of a hemispherectomy, depending on that person's age the remaining hemisphere will take over the duties of the lost one. Even in a left hemispherectomy the subject can regain linguistic ability by forming the requisite structures in the right hemisphere. This is an example of *neural plasticity*, which is the ability for neurons to make new connections over time, as a result of learning or an attempt to regain functionality after the brain has suffered damage.

On the other side of the coin from *information storage*, or memory, is *information processing*, commonly referred to in humans as *reason*, and the final lobe in the brain, the frontal lobe, is what makes it all happen. The frontal lobe is involved in executive decision-making, planning for an uncertain future, and all the other things that humans have come to pride themselves on. This includes awareness of how to act in social situations, as notoriously demonstrated by Phineas Gage, the rail worker who was partially lobotomized by a metal pole and subsequently turned into a jerk. The credit for this processing power comes from the network of 100 billion neurons, that along with structural cells called *glia* make up the brain. A single neuron is a fairly simple cell, though it is not yet fully understood. The body of the neuron is covered in receptor appendages called dendrites that receive signals from other neurons, and fires an electrochmical pulse down a long, branching appendage called an axon.

Whether a neuron fires or not is determined by a number of factors, including the time since the last time it fired and the electric potentials at its dendrites. The gap between axon and dendrite is called a synapse, and synaptic connections are strengthened the more that those pathways are utilized. This is just one of the ways the brain changes as a result from external forces, an example of neural plasticity.

Learning is one cause of neural plasticity, or more accurately learning is one way neural plasticity manifests itself. The process of repeated exposure to a novel stimulus will rewire neural pathways to incorporate this new aspect of the environment, with the overall result of a change in behavior. This is one of the abilities of the brain that give it a wide flexibility - starting with little to no history of sensory input as infants, our brains work very hard to assemble this stimulus into a cognitive model of the world. This enables past events to influence our expectations, and great effort is made to ensure that cognitive model is *coherent*, that the model fits together in such a way to reduce contradictory conclusions, and that it *corresponds* to the external world, which allows the brain to reliably predict future events and to infer happenings outside his realm of immediate experience. Note that coherence and correspondence are the main criteria in which scientific hypotheses gain validity, and that both processes can only be strengthened when opposing models are falsified. The brain is constantly performing such checks in an attempt to reduce error in the brain itself, and to do so it must constantly be engaged in the process of analyzing patterns.

Neurons will also periodically fire regardless of input, so constantly there are clusters of electrical energy that fly about the brain through pathways established by the wiring of the neurons. Learning new things involves establishing new pathways, and is a result of the electrical signals influencing the neural wiring. Dopamine, the neurotransmitter that governs the brain's reward system, has a great influence on learning - it serves as the body's own incentive for survival-ensuring behavior by identifying behaviors as good for the body and provides reinforcement for the neural impulses associated with that behavior. In general, if an experience is anomalous enough the brain will seek to integrate it within its framework of how the world operates. This has been a challenge in understanding the brain, as static models are computationally easier to implement and verify. It also creates challenges for formal models in general, as to speak of a specific neural representation of an idea is inaccurate because those neural connections are constantly in flux.

8 Language

Of all the adaptations that humans evolved as our brain functions grew - walking upright and freeing our hands, shedding our fur in favor of bodily sweat glands, and a host of others that allowed our ancestors to climb down from the trees and dominate the grasslands - language is arguably the most significant (Trask). It is language that enabled the growth of communities and cultures through shared experience, and it is language that allows us to empathize with others by offering a means of encoding mental states. Shared experience and communication are related, in that a *word* is only useful in communicating with others if the recipient can associate the word used with the denoted experience. "Denoted experience" with regards to words do not have to be of the type that evoke a *conscious* experience, like "banana" or "pain" or "travel" (lexical words); conjunctions, articles and the like all serve their purpose in the process of creating a linguistic experience through their syntactical function (functional words).

Since knowledge of our words is determined by our communicating with people who use them, and our experiences are finite, our lexicon is finite. Syntax allows humans to combine this finite number of words into phrases and sentences, allowing us to create an arbitrarily large number of possible combinations. This ability to express an infinite number of thoughts is known as *productivity*, and is a key difference between human language and animal communication. An important feature of language related to productivity is the ability for humans to create sentences and thoughts recursively by embedding clauses inside of larger ones. "I went to the baker," "I got a loaf of bread," and "the baker is bald," can all be combined into a single phrase, "I went to the baker, who is bald, and got a loaf of bread." Syntax encompasses these grammatical rules and provides a standard for well-formed statements, analogous to those found in formal languages.

Another key feature that distinguishes human language from other forms of animal communication is the *displacement* property, the ability to discuss things that are not immediately present, spatially or temporally. Displacement allows humans to ask and answer questions like *how did we get here* and *where are we going*, and questions of possibly more immediate importance to survival like *how are we going to catch a gazelle today?* Consider the last question - ecosystems evolve in a reciprocal fashion; animals with better defense mechanisms drive evolution of predators, and predators with better adaptations and instincts for catching prey will contribute to those prey animals with better defense instincts to survive. When early hominids could hold a mental image of a gazelle in their head, along with the associated traits of gazelle predatory defense, they could plan a hunt that used their shared understanding to gather food more effectively. Imagine an early plan - "You two come at the gazelle from one side, you two come from the other, then the gazelle will run toward the high grass and when it jumps down off the bluff we will be there and we will stab it with our spears." Here the displacement property allows the humans to consider the gazelle and its tendency to flee, and to understand where each fellow tribesman *will* be in the future, though all the humans are around the campfire and all the gazelle are currently over at the watering hole. What is important is the *shared* *conception* of the word in question. The shared conception is what allows for displacement, in that a mental representation must exist prior for it to be used in language.

We can also extend this ability to talk about things not directly present to the ability to speak about things that do not have a corresponding physical representation. Consider if I had never seen a yo-yo before, and someone was trying to explain to me that one was. A description may be something like "two small wooden disks, connected by a spindle through their centers, around which a string is wrapped." Several linguistic phenomena are in play here - firstly, I must know what several of the nouns are to get a sense of the *form* of the yo-yo. By knowing what *string*, *wooden*, and *disks* mean, I know what parts it is comprised of. By knowing that there is a *spindle* connecting their *centers*, I can infer the function of the spindle, its relation to the disks and the string, and a sense of how the whole thing interacts. If it were two physicists discussing the yo-yo, they could use their shared conceptions of things like *friction* and *angular momentum* and describe the yo-yo with further specificity.

Most of the words you know, however, were not defined that way. Language acquisition is an inborn function of the developing brain, and barring extreme deprivation a child will learn a language regardless if it is spoken to or not. By observing their parents and other adults, children pick up words, without conscious effort on the part of the child, at a peak rate of about one word per hour. The child will learn the language it is exposed to; adopting a child whose parents spoke a different language than you will learn your language. Similarly, a child reared in a bilingual household can learn both languages without any extra perceptible effort on the part of the child, albeit with a slight delay in production as the brain sorts out the mommy language from the daddy language. At any rate, language acquisition requires greater effort after the first few years, in the extreme case of children who grow without any language exposure. "Feral children," as they are called, do not perform as well in production and comprehension as children who grew up learning a language, sometimes to the extent of never being able to communicate whatsoever, even after extensive attempts at rehabilitation.

A striking aspect of language acquisition is the extent to which it influences our thoughts. In a study by Boroditsky, German and Spanish native speakers were asked to rate the "association" of random objects, a 1-10 scale of relatedness or similarity. Both of these languages have gendered nouns - a linguistic trait absent from English - and the assignment of gender tends to be arbitrary, leading Mark Twain to remark: "In German, a young lady has no sex, while a turnip has ... a tree is male, its buds are female, its leaves are neuter; horses are sexless, dogs are male, cats are female ... tomcats included." Boroditsky found that even though the native speakers were tested in English, both the Spanish and the German speakers tended to associate more strongly those words with the same gender in their native tongue. This tendency for language to shape, guide, or limit cognition is known as linguistic relativity (Stanford Encyclopedia of Philosophy).

If we consider the nearly converse neurological implication of linguistic relativity, that structures inherently present in the brain influence human language, then we enter territory covered by Noam Chomky in his theory of Universal Grammar. Certain grammatical features - quantity and plurality or distinguishing nouns and verbs - are so essential to the function of human communication that any natural language will develop those features (Chomsky). These features, then, are a part of the universal grammar. The idea is that through through neurological experimentation and linguistic research, it could be possible to see precisely which features are universal and which are language-specific properties. Categorization (bananas *are* fruit), causation (*if* I eat the fruit I *will be* full), and meaning, all manners of organizing human knowledge, may have such fundamental places in grammatical structure precisely because they model processes in the brain.

Of course words correspond to certain clusters of nerves, but more importantly *meaning* in those words is found in the neuronal connections. The word "banana" is forever tied to parts of the brain for the color yellow and the smell of the banana, but also by association a cascade of other types of fruit, a mental image of sunny tropical banana plantations, pyjamas, the word "band", and a whole host of other firing patterns that ignite with varying strength, depending on the level of connectivity. By using words, or retrieving meaning, those pathways are strengthened, and by relying on similar connections in the audience, someone using language actually invokes those connections in the mind of the listener. The semantic meaning of the word or symbol is not within the word or symbol itself, but imbedded in the network formed by our associations, which itself is dictated by the innate grammar of our understanding. Grammar generates meaningful phrases, but it is also the reason for meaning itself. Noting this, Charles Henry writes, "that the neurological arrangement of the brain is also fractal in nature is no coincidence. In this regard it is impossible to separate the dancer from the dance: to see grammar as an abstraction removed from meaning."

Grammar, then, is the way of connecting concepts in a meaningful way - that certain words put together in a particular way means the audience can extract some information from it. In the language of formal languages, such statements are called well-formed. However, natural language has the further test of veracity - to what extent does the statement correspond to the outside world? As we have seen, no symbol-system completely conforms to reality, so any statement cannot conform 100% to reality, but some statements certainly do more than others. "Hitler started World War II" has some truth to it - without him, World War II certainly may not have happened, but all manner of influences political, economic, and social made it possible as well, many of which had nothing to do with Hitler. The phrase “Ryan Seacrest started World War II,” however, has no correspondence to what happened, namely because its impossible to be an agent of causation when you don’t even exist yet, but also because they would not let him on television after doing something as heinous as that. What is important is that the rules of formation in natural language do not act on the meaning of the words being strung together, but on their grammatical function. It is up to the audience, then, to deconstruct the sentence and evaluate it for itself. But if indeed grammar is to some extent innately wired in humans, then this process would seemingly be instantaneous; the grammatical deconstruction of “Ryan Seacrest started World War II” immediately pins Ryan Seacrest as a causal agent, invoking the deeply wired structures that comprehend causation, Ryan Seacrest, and the causes of World war II, yielding a deep seated disinclination to evaluate the statement as true.

We could make the argument that the degree of correspondence between what we say and what is true carries within its relation an inherent truth value, and if we were to keep the analogy between formal and natural languages then true statements about the world can in some sense be reduced to deductions from axiomatic facts. The key difference, however, which separates the two is that natural language completely lacks “rules.” Yes, syntax is critical for forming sentences others can understand, and truly there are many books written on English grammar, but the only metric for making, and evaluating the truth of, statements in natural language is that other people can understand them and evaluate them as true or not. So, beyond any inherent universal grammar, the rules of language depend on those with whom you are communicating to have a shared enough experience to understand the language you are using. The particular way a person speaks their mother tongue (first learned language, L1 in fancy linguistic abbreviations) is defined by their interactions with other speakers, and the rules that they follow when producing sentences are the only set of rules that truly matter.

Even teasing apart the neural connections of which syntactical and lexical language production in a formal manner is difficult, as the process is highly dependent on the brain state, context, and probabilistic interactions between the neurons themselves. That is one reason for certain errors in communication - that sometimes a construction in our head that sounds reasonable throws another person off guard, perhaps because they were expecting a more familiar construction but also because the rules are arbitrary. In a formal system, or at least the context of the formal system, effort is often made to establish systems where one statement cannot mean two things, because systems like that have a number of nice properties, especially regarding formal study. However, in natural language, such ambiguity is common or even taken to be a part of the language itself - languages like Japanese hold as much meaning in the context of their interactions as they do in the interactions themselves. Tone of voice, facial expression and even the level of pupil dilation all assist in conveying our thoughts but they are usually not counted as part of grammar.

While tone of voice or facial expression are usually *not* part of grammar, there is certainly nothing stopping languages from incorporating them as a part of grammar. American Sign Language (ASL) for one has a language that consists of nothing *but* gestures and facial expressions. Since gesturing is markedly slower than speaking, in order to keep the same information density many verbs are implied, like "to be" and the ubiquitous interrogative "do" are often left out, and body position conveys other information. In a recent example, New York mayor Michael Bloomberg went on television before Hurricane Sandy to brief New Yorkers on the coming storm. At his right was interpreter Lydia Callis, who became an internet sensation after highlights of her energetic and emotional performance made their rounds on youtube. Her facial expressions seemed big and dramatic, but this is important for denoting certain grammatical traits like quotes or hypothetical situations (Mitchell et al.). In addition, grammar is spatial in ASL - moving left to right is not a rhetorical maneuver in ASL, rather when one is speaking (or signing, more accurately) they can establish certain "zones" in areas in front of them - middle left, top right, and so on - which refer to different subjects they are signing about. It's as if the "spaces" are pronouns, which can be assigned and moved on the fly and good ASL speakers can keep up with all 9 (Cormier). Even though this language is so different in its form, not even being a spoken language, it still involves the Broca's and Wernicke's areas just the same - brain scans of subjects signing and speaking when identifying objects in pictures were nearly identical (Emmorey et al.).

ASL, and signed languages in general, strike an interesting balance between natural and constructed language. In ancient societies where deafness occurred, children born deaf would communicate in some way with the people around them - without auditory input to process, understanding the people around him meant observing what they did with their bodies, in terms of gestures or expression. We could expect early signed languages to be pidgins between the hearing and the not, but as education and interest concerning deafness grew in the 18th century, so did efforts to allow the deaf to communicate with others, and rules and signs became standardized. Thus the natural component of language met with a little bit of planning with regards to vocabulary and grammar, and ASL was born. There are a number of reasons to plan a language, however, and they need not be practical. It is well known, for instance, that J.R.R Tolkien was a linguistics nerd first and an author second - *The Lord of the Rings* trilogy served as a vehicle for the languages he created, like Elvish, Dwarvish and the black speech of Mordor. Fans of *Star Trek* are familiar with certain Klingon phrases or even profess fluency in the language, but the only reason Klingon exists is to give a fictional universe depth.

More interesting to linguists are a subset of constructed languages called planned languages, which are intended for human use. Esperanto is arguably the most well-known of the *international auxiliary languages*, designed to facilitate communication between people across natural language barriers. Its inspiration draws heavily from Latin-based and other European languages in both vocabulary and syntax, however, and its bias in this manner is one of the obstacles to its goal (Cowan). Lojban is a more ambitious project, which takes its vocabulary from Mandarin, Hindi, Russian, English, Spanish, and Arabic, and bases its syntax and morphology on predicate logic. Because of this, each written or spoken statement in Lojban has precisely one semantic meaning, removing ambiguity by precisely defining and classifying words with regards to their function in a sentence, as well as the relations between words having only one meaning. The result is a language that is unambiguous, even to computers, and there are great hopes for using it as an interface language between computers and humans, or an ideal intermediate language for use in machine translation.

Lojban is truly odd in several regards. Firstly, no natural language can boast of semantic unambiguity, and while it is not normally a problem based on context, constructions can be made in any language that exploit that reliance on context. A politician might say "I oppose taxes which hinder economic growth," and what exactly is hindering the economic growth - all taxes, a certain subset of them - is unclear, and linguistically the politician hasn't lied per se, but also hasn't really said anything. Second is the seeming reversal of the role of culture with regards to language. Language is a defining aspect of groups of people - it marks the difference between people with whom you can communicate and whom you cannot. Further than that language and culture, at least in the years before widespread global cultural and linguistic homogenization that has left us with McDonald's in Java, are defined in terms of each other. Groups in New Guinea who split from other tribes make efforts to obfuscate their language such that in a couple of generations the two languages are unintelligible to the other. Jokes especially rely on references to shared experience or exploitations of thought processes that are influenced by language, and the resulting unifying power of laughter strengthens community. Lojban exists completely outside of that relationship - its cultural heritage comes from all over the world, and the properties of natural language that allow for certain varieties of wordplay and idiom are removed.

If the purpose of a language is to share ideas and relations between objects in the external world, then Lojban is spot in its design. The way language evolved in our brains, however, is quite imprecise. Language relies on the connectivity of words to produce meaning, and its flexibility is what allows for its expressiveness. We are constantly seeking connections because that's how our brain is wired - patterns in visual stimulus are aggregated into a whole, associated with sounds or feeling, and are mapped onto our mind's eye as a part of the outside world.

Noticing patterns in the systems all around us give us insight into the rules that govern them. Mapping, associating, and importantly analogizing are what gives our cognitive edge over other animals, and this property is thusly reflected in our language. Metaphor is the process of drawing connections, and language relies on this - just in the realm of animals, we can say in english "sly as a fox," "wolf down," and "chicken out," and because of our shared experience we can get the idea of what those phrases mean without any prior knowledge of them. So it is with every aspect of our thought, as beyond there really being a soul, or stuff separate from matter, all our experience is reducible to neural firing patterns. That is not to say that Lojban is without expressive language, or humor or poetry, but by removing a dimension of connectivity - logical ambiguity - there may be a cost in expressivity. This remains to be seen, as only a handful of people are fluent in Lojban, and even fewer (perhaps zero) have an L1 of Lojban. This is perhaps its most interesting application - if we could have evolved a logical language, why didn't we?

The current state of Artificial Intelligence with regards to understanding and communicating in natural language is nowhere near mature, but from what we know about language and human intelligence, the two are deeply intertwined. Firstly, something in the brain functioning as an internal grammar seems to be the driving force behind not only communication with others but also within our own minds. Thus an effective epistemological system within a computer brain would take on the form of some sort of language that could describe relations between concepts represented by words. However, too much of a focus on individual pieces of knowledge would then detract from a more rich understanding that is based on the connections between concepts not as pieces of knowledge but as results of experiences and interactions with the outside world and with other agents acting therein.

9 A Brief History of AI

In 1955, a group of researchers from across the Northeast working at places like Harvard and Bell Laboratories proposed a conference the following summer, to be held at Dartmouth college, to investigate certain problems in artificial intelligence. The 10 member team would work together over the course of two months working out the kinks in machine learning, natural language processing, creative problem solving, and other foundational problems in AI (Minsky et al.). According to the proposal they believed that "a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer." It seems that the Dartmouth scientists were not quite aware what they were getting themselves into. This is understandable, as no real effort had been made in the field before the conference, and the topics treated in the proposal are all still active areas of AI research over 50 years later.

The Dartmouth conference is considered the birth of the true search for AI, but the idea of a human creation bestowed with intelligence is old as time, and much of the groundwork for artificial intelligence had been laid for centuries. Developments in the fields of logic and mathematics, such as Aristotle's syllogistic reasoning (A implies B, B implies C, so A implies C) and work on formal systems have influenced reflections of our own thinking, and have given us inspiration for the direction of logical faculties in computers. Corresponding developments in technology, namely faster computers and high-level programming languages, made a "thinking machine" a definite possibility.

What constitutes "thought," however? What makes a "thinking" machine different from a "calculating" one? A long standing argument to the feasibility of AI is the assertion of the existence of a "soul" or some other immaterial property of mind, but as it seems human beings have conscious activity located in a physical network of deterministically operating neurons (as well as that this assertion is antithetical to the central concepts of this thesis) , we will disregard dualist discussion and will choose a more reasonable definition of intelligence. Human intelligence is a matter of forming a network of symbolic representations that correspond to the outside world, allowing us to act within and make decisions about it. We convey these symbol-systems through natural language, which allow us to correct our own conceptions while forming a theory of mind that allows us to recognize a similar intelligent process occurring in the heads of others. We are aware of our own intelligence, and verify it in others through interaction and communication.

It is on this basis that Alan Turing, in a 1950 article, proposed a test for artificial intelligence now known as the Turing test. One human, the interrogator, sits at one computer terminal engaged in two chat sessions (in 1950 it was called "teletype." How quaint!), with a human on the other end of one session and a computer at the end of the other. The interrogator asks both entities questions for an allotted period of time, at the end of which the interrogator must decide which is man and which machine. So not only must a machine possess a large amount of knowledge on a wide variety of subjects (common knowledge), it must also be able to make statements about them that line up with rational human thought (common sense), as well as the ability to attempt tasks given by the interrogator. The machine may be asked to play checkers or write poems, and must answer as a human would if it had any hope of passing the test. The Turing test is a way to judge intelligence independent of how it is housed and without cheapening our definition of "intelligent" by explicitly defining specific tasks we would expect an AI to be able to accomplish.

It is with the concept of a Turing test in mind that the Dartmouth conference set their goals for AI, but the early years in AI saw little progress toward a machine that could pass it. Instead, initial research in the field was focused on finding human-style solutions to problems by combing search trees with the assistance of heuristics to lessen the computational burden caused by the exponentially increasing number of trees (Mainzer). To see what I mean, consider a computer trying to play a game of chess. The starting move alone includes 16 possibilities for pawn movement and 4 for knights, yielding a total of 20 moves to consider in the first turn alone. All those possible moves leave the board in different positions as well, so the next move will include different possibilities if the king's pawn had been moved as opposed to the bishop's, and so on. This arrangement of possibilities that branch off as the game progresses are search trees. A heuristic may knock out many of these possibilities - for instance, any move tree that would include losing a piece might mean that its children are ignored, for generally in chess more pieces means a greater advantage. Discouraging opening moves with pawns on the edges, playing for control of the center, and many other elements of chess-playing strategy are included in these programs. Deep Blue, the first computer to beat a sitting grandmaster, also included a database of opening and closing positions, lessening the search load, and regularly looked 5 or 6 moves into the future but also sometimes as far as 10 moves ahead. Deep blue also had the benefit of 30 separate processors, having its code written in C, and processing an impressive 200 million positions per second (Hsu).

Problems from this era were not limited to chess and checkers; any problem that can be sufficiently formalized can be investigated in the manner. The General Problem Solver, created in 1959, was designed to do just that - generally, to solve any problem that was sufficiently formalized (Simon et al.). In this case, formalized means that the problem could be coded in the program's system of objects and operators. GPS's heuristics were straightforward, driven by the "means-end" method of reducing differences in a current state and a goal state and a recursive problem solving method that broke complex goal states into groups of less complex goal states. This approach was based off a psychological study wherein the subjects said aloud their problem solving techniques, which the creators of GPS refined and formalized. It was able to prove the first 38 theorems in the *Principia Mathematica* and some trigonometric identities, as well as play a little checkers. Anything much more complicated than that and the search space exploded and the program became impractical.

The late 60's saw a shift in focus from defining and refining search patterns for specific problems to one that was more concerned with knowledge representation and symbolic manipulation. This meant for a burgeoning interest in natural language as a result, as language not only offered an interface with the AI (Winograd) but showed the ability of the knowledge representation. The quintessential example from this area is known as SHRDLU. A user sitting at a terminal could ask SHRDLU about its world, which consisted of blocks of various colors and shapes that could be stacked, moved around, or moved into other containers. SHRDLU would respond, clarifying instructions or telling the user about certain blocks. A sample dialog from a famous demonstration follows:

Person: Pick up a big red block.
Computer: OK.
Person: Grasp the pyramid.
Computer: I don't understand which pyramid you mean.
Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.
Computer: By "it", I assume you mean the block which is taller than the one I am holding.
Computer: OK.
Person: What does the box contain?
Computer: The blue pyramid and the blue block.
Person: What is the pyramid supported by?
Computer: The box.
Person: How many blocks are not in the box?
Computer: Four of them.
Person: Is at least one of them narrower than the one which I told you to pick up?
Computer: Yes, the red cube.

SHRDLU has 2 main features that allow it to operate inside its world and interact with the user. First, SHRDLU lives in a simple universe with just a few shapes, like "block" and "cube," and descriptors, "blue" and "taller" and so on, such that its entire vocabulary can be contained in as few as 50 words. This made the relations between the words easier to parse, hence easier for SHRDLU to "understand." Secondly, SHRDLU has a basic memory. that it uses for context for interactions with the user. The user asks "what is the pyramid supported by," and since the program just talked about the blue pyramid it can infer that the blue pyramid is the one the user is referring to. Similarly when there is no context for "pyramid" and the user asks SHRDLU to grab *the* pyramid, SHRDLU doesn't do anything because it knows "the" indicates a specific object. The memory and language together allow it to answer questions from the user, and to remember new words it learns through their interactions.

It was around this point, however, that AI research began to slow down, mostly due to constraints on space and processing power (Mainzer). Natural language AIs, for instance, need a large amount of information about the world - a database of common sense - and the memory of computers in the early 70's simply wasn't enough. This led to a withdrawal of funding from agencies who had been disappointed by lack of progress, and the research community stagnated. It wasn't until the early 80's that widespread adoption of computer systems in businesses and a new kind of AI - the expert system - rose in popularity. An expert system is designed to mimic the qualities of a human expert in and function but not form; using a database of specific knowledge (and avoiding the problem of common sense), the expert makes a series of deductions about the system, creating new knowledge or asking for more information from the user. The knowledge base is separate from the inference engine and designed to be programmed somewhat naturally, using simple IF - THEN statements, allowing for one inference engine to be used in a variety of different applications, like medical diagnosis or process control.

One famous such system is called MYCIN (Buchannan et al.), created at Stanford and designed to diagnose and treat bacterial infections of the blood. Just like a medical expert, or doctor, it will ask questions, order tests, synthesize the results into a diagnosis and order a treatment. One of its rules for diagnosis is

IF

1. The gram stain of the organism is gramneg AND

2. The morphology of the organism is rod AND

3. The aerobicity of the organism is anaerobic

THEN there is suggestive evidence (.6) that the identity of the organism is bacteroides.

MYCIN also has similar rules for diagnosis:

IF:

1. The therapy under consideration is one of : cephalothin clindamycin vancomycin AND

2. Meningitis is an infectious disease diagnosis for the patient

THEN the therapy under consideration is not a potential one.

MYCIN consistently outperforms medical students and general practitioners, but it has yet to see use in the public sphere due to liability reasons (Ginsberg). If, say, a bug in the software occurs and a patient dies, is the doctor responsible? Is Stanford university? There are moral and ethical objections to AI and they come from many different places, but in areas without life-or-death consequences like business and engineering expert systems have found diverse applications. By 1988 there were 139 of them implemented and in use as listed by Feigenbaum et all.

Since then Artificial Intelligences, or at least technologies that were developed as a result of research into AI, have matured to the point where they see daily use in computer systems. Tools like expert systems and knowledge representation have seen a migration from the AI lab and found their way into the standard toolbox of programmers. Unfortunately for AI researchers, once a problem is better understood and AIs can reliably complete them, those programs are then detracted as not being intelligent and the problem reconsidered as one not needing intelligence to solve. This trend will end in one of two foreseeable ways. Either AIs will get to the point that they exhibit behaviors as truly intelligent, creatively solving problems, composing symphonies, and talking naturally, or we will reevaluate what special qualities define intelligence and discover that we may not be all that special ourselves, and that our own mental process are mechanistic to a profound extent. Those two possibilities are not exclusive.

10 Neural Networks and Neuromorphic Hardware

Efforts in creating Artificial Intelligences draw heavily from models of human intelligence, for several reasons. First and foremost is it is much easier to draw inspiration from nature than start from scratch. Because nature has offered us nothing (so far, not to discount the possibility of extraterrestrial minds) besides our own brains to model cognition, feeling, and even some of the lower-level neurological functions, our own minds are the prime target of study and simultaneously the things doing the studying. This leads into the next point, that we are setting the standards of what is "true" AI by the capabilities of human brains. Much of the successes in AI have been modeling high-level attributes of human intelligence, like use of heuristics in search and knowledge representation, but these tend to be models for what is actually an aggregate of many levels and stages of processing over a wide variety of specialized structures.

The problem of more accurately modeling and simulating a human intelligence is twofold: first, we need to know more about brains and nervous systems and how they are all wired together so that we have a better idea of what we are emulating, and second, we need to figure out ways to implement the kind of computation these models imply. The traditional model of a practical computer, the model which all modern computers follow, is the Von Neumann machine. Named for Jon Von Neumann, yeat another famous Hungarian mathematician who contributed greatly to anything in mathematics that was a hot topic in the first half of the 20th century, the Von Neumann machine abstracts and relates the basic parts of a computer. A control unit retrieves instructions from programs and orchestrates the computations carried out in the ALU, or arithmetic-logical unit. Included in this description are memory for use by the program, known in computers today as RAM, and mass storage, known as a hard drive.

This method of computation differs from natural neuronal ones, and its differences stem from features that made this model an attractive one for building computers. The Von Neumann machine uses discrete units which have specific purposes, and if we know anything about the brain, nothing within it is truly discrete and no operation is centralized in any specific part. This is especially true of memory and reasoning - each is facilitated in the brain by the same neural substrate. The second is that the specific flow of operations - grab instruction, send it to the ALU, grab the contents of relevant memory addresses, perform the operation, output the result, and clear the registers - is inherently serial. The brain performs parallel processing such as in the visual system, where part of the system decodes shape, another color, and so on, each processing their respective attributes simultaneously and assembling the raw data into a whole. No single area of the brain is ever the only one active, and computationally reasonable schemes for modeling consciousness will likely need to reflect this. Both of these problems admit essentially the same solution set - model this connectionist structure and function in data structures, and represent them as abstractions over the current type of hardware we are using now, or come up with a new hardware model that inherently exhibits connective, parallel approaches to computation and storage. More compactly, do we implement the neural metaphor at the software or the hardware level?

Implementing neurons as data structures is the idea behind neural networks (hereafter NN), which emulate the functionally important behavior of neurons, namely the tendency to fire when the potential at the synapses exceeds a certain weight. Each artificial neuron, which I will call a node, is governed by a simple function that essentially sums the inputs fed to it from the nodes connecting to it. These inputs are weighted, and the weights are determined by a learning process, which take a set of data, perform the calculations over the set of data, and make notes of successes and failures and modify their weights accordingly (Zurada). The simplest kind of neural network is the feedforward variety, wherein nodes only provide input data to nodes further along in the computation (connections are never backward, and hence are free from loops). This is the motivation for the idea of layers in a NN, with the first layer that accepts the input data for the whole computation called the input layer, hidden layers for performing the required computation and a dedicated layer for relaying output.

What is essentially happening at each node in each step of the calculation is inputs to the neuron x are decided by performing a calculation on thee weighted input $w_1\ y_1$, $w_2\ y_2$... and outputs a single number, so it can essentially be thought of as a vector-to-scalar operation. This is known as the activation function, and in its simplest form is binary - if the inputs exceed a certain threshold, then the node activates and fires. A single-layer array of these, called a perceptron, can function as approximators of certain functions, including logical connectives like AND and OR, but their robustness was shown to be limited,

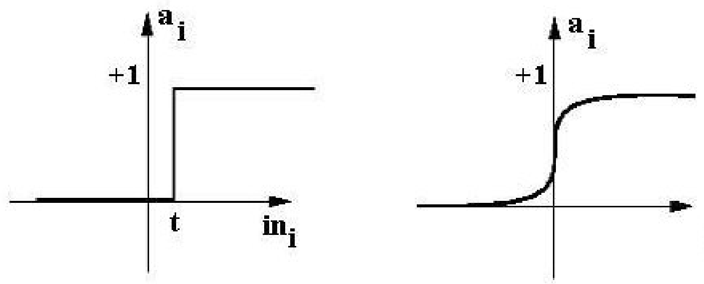

Figure 5: Left, the binary step function. Right, the sigmoid function

notably in the inability of perceptrons to model the XOR, or exclusive OR, relation. So the binary approach not only a naïve view of the operation of neurons and an ineffective activation function. More appropriate is the (logarithmic) sigmoid function, given by $S(t) = \frac{1}{1+e^{-t}}$, which more closely models the actual activation function of a neuron. It resembles the threshold function, so the basic functionality is the same, but because of the "wiggle room" near the extremes of the graph and the fact that it has a continuous derivative, its statistical approximating power is greatly enhanced. The derivative also lends itself nicely to several methods of learning, such as the steepest ascent method, where the error E is represented as the difference between the target value and the sum of the weights in the layer before it, or the actual computed value. The weights are treated like a vector, and adjusted individually in the direction of the negative gradient of that vector, which is easily computed thanks to some nice properties of the sigmoid function.

The computational power of NNs lies in the connectedness of the nodes. This is the main human influence in their design, and beyond the specific node-level decisions like picking good activation and weighting functions and an effective learning scheme, the way said nodes are networked is the most important part of creating an effective NN. This is still something of an art, as networking them together relies on knowledge of the problem to be solved. Nonetheless, NNs have been implemented successfully to solve a wide variety of problems in a number of settings, especially in situations where there are lots of dependent factors, each having some probabilistic influence on the outcome. Cancer diagnosis, for instance, relies on a number of different test results, patient histories, and other factors, and NNs have been shown to be effective in diagnosing colon cancers successfully over 90% of the time. If cost of tests and procedures is known, they can also predict cost the cost, and they can predict survival rates over a 5-year span as accurately as 95% (compared to 80% for the physician rate the data was measured against, Ahmed). Additional uses have been proposed for predicting stock market change, facial recognition, and image processing (Ginsberg).

A particularly interesting use of NNs in pattern recognition was demonstrated by the network NETtalk (Sejnowski). It is designed to learn how to pronounce letters given the letters surrounding them as context, with the hopes of having the network learn how to read text aloud. Traditional approaches to having computers read text look up the words they "read" in a dictionary, which contains a phonological breakdown of each word, assemble those phonemes, and synthesize them. NETtalk, on the other hand, contains an input layer of seven groups of 29 nodes, which correspond to seven letters - the middle is the letter whose pronunciation is in question, and the remaining six represent the three letters (or spaces or punctuation, hence 29 nodes per group) on either side that serve as context. Its output layer contains 26 nodes corresponding to 23 features of phonetic production like nasality and dentrality of consonants or the open and closeness of vowels that generate the International Phonetic Alphabet, which is essentially all the sounds a human mouth can make, along with three other nodes for encoding boundaries in syllables and stresses. From these output qualities each letter is synthesized as the input is read, as well as gaps between words and sentences that manifest themselves as resembling human speech. When trained on the transcribed speech of a first grader, the accuracy of its reading gets markedly better after just a few hundred runs, and the progress the NN exhibits parallels infant speech development but at a much faster rate. What starts out as computer babbling quickly takes on a phonetic form, and little by little the larger patterns are recognized and learned, and the final weights in the learning process yield recognizable speech. While effort was needed to train the NN, the end size was much smaller than the contents of a whole dictionary and its associated rules for formation, making it much more space-efficient.

The success of neural networks in applications of conditional probability underlie several theories about human nervous processing. It has been suggested that much of the brain's processing, from associating and assembling sense data to higher-level decision making, is done probabilistically (Specht). Through the immense amount of experience and learning a human being undergoes, the brain gains an appreciation for its tendency for error, and takes this into account when mapping acoustic data to visual data (Knill). Establishing where an object is likely to be given noise and error in sensory input is a classic setup for a Bayesian analysis, and not only is such a problem well-suited to a NN solution, there is also evidence that the brain itself, and not just the human brain, utilizes this method which has barely been known to mathematics for 300 years but apparently had been figured out by evolution some millions of years before that. Probabilistic reasoning is not alone - there is evidence that the basic levels of the auditory and visual systems use some level of Fourier decomposition in encoding data for the rest of the systems (Graham), possibly to lighten computational or sampling load. The possibility of such a discrete analysis to be implemented with neural networks has also been shown (Velik).

Two main problems persist in the study of NNs, and they parallel the two main problems in AI in general. The first is that while more layers of neurons enable NNs to solve more complex problems and solve certain problems more efficiently, they become equally more demanding with regards to complexity and efficiency in design. Larger amounts of connections and a wider variety of training data are needed for real-world applications, which model the need for large databases of "common sense" and reasoning abilities in rule-based knowledge systems. All of this complexity leads into the second problem, which is the lack of computational power we have to run large networks. Though the design is connectionist, the hardware that NNs run on is still serialized, and the combinatorial tendency for the number of synapses to exceed the number of nodes by several orders of magnitude means that a traditional computer still has to sort out huge numbers of calculations to drive the network. Moore's law, really a rule of thumb coined by Intel co-founder Gordon Moore, is often cited as hope for this predicament. It states that the number of transistors in a processor doubles roughly every two years, which means that implementing larger NNs is simply a matter of time. Moore's law may be reaching its limit, however, as the channels for electrons get smaller and smaller, approaching the length where it is entirely likely an electron will tunnel over and cause an error (Zhirnof et al.). Different paradigms of computation may be needed entirely, and if AI research doesn't drive it then consumer demand certainly will.

With that in mind we should take a look at what we already have with enough computational power to implement large networks. The first thing that should spring to mind is *the* network, the information-processing information superhighway known and loved as the internet. It is not, at any point in the near future, going to become self-aware and able to answer any question we ask of it, mainly because at this point it is simply a network of computers sharing information with one another instead of an organized problem-solving, knowledge-organizing machine. This is a result of the fact that nearly everyone using the internet is running software to do the former and not the latter. However, the information-sharing network is there, and its development in recent years and for the visible future has been an interesting one. Far gone of the days of carefully crafted HTML documents with pictures of dancing hamsters and whatever other content contained therein - personal websites are now in the realm of blogs and social media pages, and networks of not only people but ideas have become the main focus of the internet, with an emphasis on the separation of form and content. Tags that describe content of pages or pages themselves are standard, whether visible to users or not. This move towards having websites being containers of content, with user-generated descriptions thereof, marked the start of "Web 2.0," though everyone still calls it the web, and the delineation is really only useful when we start thinking about what

“Web 3.0” will look like and what it looks like already. A driving force in this development are trends toward a “semantic web,” one in which tags not only provide context for humans but also for machines to understand the nature of said content and how it is connected with other data on the internet. Such a standard would allow computers to use all the information available on the internet in making decisions, linking markets and trends globally.

This is all contingent on the maturation and adoption of several technologies for representing and transmitting data in a machine-readable way, so the realization of the internet as a global AI remains to be seen. One entity that may pick up the slack is Google, who not only have a track record for supporting such standards, but also have a stake in such a web. Knowing how to locate exactly what the user wants is half of Google’s business model, and tagging content is essential to organizing the data it stores. The other half is knowing what the user wants to search for - the semantic meaning of the input itself, and Google maintains an active AI research lab to facilitate that. Realizing the search engine as an AI is one of the long term goals of Google, quoted by Larry and Sergei themselves. The lab recently published a paper about neural networks that were not only layered in their implementation but also *deep* - up to 8 layers so - that could recognize faces even when the input was not labelled (Le et al.). They also clearly have an interest in language processing - seen by their interest in projects like voice search on android phones and Google Translate. Ideally, a search engine could pass a Turing test, as it would know exactly what you meant based on its knowledge of you with regards to its past interactions with you, knowledge of where you live, and most importantly it could parse your input naturally and figure out what you mean semantically, and return the perfect result to you. With their huge computational, creative, and financial resources, it’s entirely likely that the first AI could be born not at MIT or the Department of Defense but at Google, and in hopes of avoiding a Terminator-filled future then we will truly see how deep the “don’t be evil” motto at Google runs.

Taking the idea of modeling neurons *in silica* and grinning and bearing the computational load has been taken to its literal extreme by the folks at IBM. Instead of a running a NN, which heavily abstracts the neuron metaphor, they run a program called NEURON, a research tool designed to accurately model the structure and behavior of actual neurons and networks of neurons in features like dendrite arrangement, axon length, and neurite branching. The program, called Blue Brain for the supercomputers that IBM donated (Markram), has shown success in completing its original goal of modeling a rat’s cortical column - a functional group of neurons that span the six layers of the neocortex. In humans there are some 60,000 neurons in such a column, but in rats it is substantially less, closer to 10,000. The project mostly exists now to test new research in neurology, and the group has identified some 50 different varieties of specialized neurons. They have an eventual goal of running a simulation of a complete human brain, but for all the reasons already stated this is a far off goal - while success with a 10,000 neuron arrangement is laudable, its a paltry portion of the 100 billion or so neurons in a human brain. Modeling neurons with such accuracy adds a burden to the computational load, requiring a whole abstract representation of a neuron in the functional bits of a computer. Biological nervous systems have the advantage in this regard, as the neurons are themselves the functional bits.

So it remains that current methods of computation remain as a key factor inhibiting deeper research into AI. The parallel, statistical model of neural computation has proven itself in theory via artificial neural networks, and from what we have seen with regards to the structure and function of the brain. In animals it provides a manner of computation in which *hardware* and *software* are one in the same, programs without a programmer. It is on this front that the next developments in hardware will take place, and it seems there is no feasible way to mimic the computational power of the human brain using traditional, serial, instruction-based computers without going big. This parallelism and connectivity would be reflected in the hardware, and continuing in the successful tradition of mimicking human intelligence would likely take the form of an artificial nervous system. This approach to engineering is a *neuromorphic* one, and it could be realized in several different ways.

Several different research labs, including IBM and HP, are involved investigating neuromorphic designs for computing, but the main player has been DARPA, the same agency responsible for the network ARPANET, which would grow into the internet, and a trillion dollar (expected cost) fleet of unstoppable joint strike fighter planes. The SyNAPSE project, which stands for Systems of Neuromorphic Adaptive Plastic Scalable Electronics, has already spent some 50 million dollars in research grants in hopes of building such a computer (Lohr). IBM and HP were the main recipients of the funding, and each have produced chips that mimic neural activity in hardware using slightly different approaches. IBM's prototype is dubbed the *neurosynaptic core*, and represents axon-dendrite connections in a matrix of synapses. This "crossbar" model allows for memory management and computation to be done simultaneously, avoiding the bottleneck in input/output channels that is characteristic of the Von Neumann approach (Merolla). Computation is done as the processor receives requests from firing synapses, which reduces power consumption by making computation on-demand as opposed to clock-driven. Network architecture is specified before the machine is run, and the team at IBM reported a 94% success rate at identifying written letters as well as other successes with NNs performing other tasks, with hopes of creating a dynamic system that can make decisions based on environmental factors.

HP's offering is radically different, and it stems from a breakthrough at HP labs just a few months before the SyNAPSE program was announced. In 2008, HP demonstrated the world's first memristor, predicted in 1976 but unrealized until then, which is the fourth "fundamental circuit element," along with the resistor, capacitor, and inductor (Marks). A memristor functions much like a resistor, but its resistance changes with the magnitude and direction of current flowing through it. When current stops, the memristor's state remains the same. Furthermore, thanks to the design of the memristor, which is essentially a layer of wires with an oxide channel angled perpendicular to the wires, they can be incredibly small and function as memory without much heat output and without constant electrical power (Snider). While still far from mature, including issues with unreliable production techniques and a tendency for error, memristor technology will hopefully in the future enable us to integrate computational elements, working memory, and mass storage aspects of the Von Neumann model in a small, energy-efficient and low heat output form factor. All of these qualities address the traditional problems of implementing parallel, neural-based methods of computation.

MoNETA is the result of a memristor-based approach to cognitive computing. It models three important aspects of neural computation using specialized approaches to each. First, to model the huge linear algebra problem that is calculating electric potentials across a neuron's possibly hundreds of dendrites, MoNETA uses a "dendritic" core that is essentially a GPU, or graphics processing unit, with a memristor bank serving as memory (Versace). GPUs are specialized parallel processors normally used inside graphics cards to perform a small scope of problems that make up the computational bulk of graphics processing, linear algebra problems that are associated with rotations and translations of the matrices that represent objects in the digital world. The decision within a neuron to fire or not is handled by a CPU-like core, which is flexible enough to handle a wide array of instructions that model activation and transmission functions of the neuron. Finally, the axon-dendrite interface data will take to form of a memristive memory array. Each unit of these will be a single core, and like the team at IBM, the future for MoNETA is a matter of making several of these cores all network together, scale, and eventually be wired together to produce an artificial mammalian intelligence, called an animat, which will navigate, learn from, and survive in an information-rich environment. The MoNETA team anticipate such a brain with the computational power of a rat's will fit inside a shoebox, quite a lofty goal considering the Blue Brain can only model a single cortical column, and its hardware fills up a whole room.

One day every piece of consumer electronics might possess such a brain. With the limitations on transistor size already influencing Moore's law such a development will not only have practical advantages but a shift to parallel, pattern-based computing is going to be necessary to move the computational power of computers forward. The realization of these goals for neuromorphic hardware is still far off in the future, both because these technologies need to mature and a more complete picture of how the brain works is needed. The first is simply a matter of time and funding, and with the successes demonstrated so far it is likely that funding for these sorts of projects will continue. The second has been the "big problem" in neuroscience since the brain was first studied, and it certainly seems like the biggest obstacle to realizing AI. There reason for hope on this front, and it is precisely all this research being done into neural computing and computational neuroscience that I've been discussing. Computer modeling of the brain, and the ability to test quickly new hypotheses in neuroscience will accelerate discoveries about the brain and bring us closer to a fuller picture of how the brain works.

11 The Future of AI

Even if we never get around to creating an AI, the chance of which my Bayesian brain has calculated to be on the order of a snowball's chance in hell, a full picture of the brain would be invaluable in improving the quality of life for everyone. First and foremost, with a detailed picture of the brain psychologists could finally quit hearing about how psychology isn't a "real" or "matured" scientific discipline. Secondly, with an explanation linking form and function, mental illness could be diagnosed and treated much more effectively. If schizophrenia, bipolar disorder, and sociopathy were a thing of the past then we could elevate the humans who sit at the lowest rung of society, and bring society as a whole forward. Still more exciting is the possibility of brain-computer interfaces. Nanobots that sit in your glia and monitor neural impulses in various spots of the brain could provide a completely integrated interface for computing. Imagine wondering what the capital of Botswana is and instead of looking it up, you literally just look upward and download the answer into your brain. Understanding it as a machine would mean its just another machine to hack - the possibility of a superhuman intelligence coming in reach of humans is very real. Finally, with the appropriate hardware, what's stopping the information-system that is your mind from existing in a different substrate? The ability to upload and transmit our consciousness itself is particularly exciting because it effectively means that "you" - not your fleshy husk or the wetware you run on, but "you" as an individual mind - could be immortal. This is the holy grail of technological achievement, as technology exists to make life easier. With no physical needs and no risk of dying at any point in the near future, life would literally be a no-brainer.

Assuming it does happen, what does the future of AI look like? A growing movement led by futurists like Ray Kurzweil anticipate that not only will we build computer that will pass the Turing test, but we will build an AI with cognitive and computational abilities that far outsrip those of humans. This means that not only will it be better at math, managing resources and coordinating markets, but it will be better than humans at *everything* we use our brains for. A well-designed AI would could govern better, more objectively, and without the pesky love of money that has plagued our political systems since there was money and political systems for it to plague. Most importantly a well-designed AI could design AIs better than humans could. The moment this happens, more and more powerful AIs spring up, and the rate of computational increase would explode. Commonly referred to as the singularity, it could prove to be the moment that would mark the end of human life as a struggle to live within its environment and begin a new era of art, peace, and free expression. Alternatively, it could mean the enslavement and inevitable extinction of the human race, and mark a future filled with Terminators and buggy HALs that send us flying into the black abyss of space.

Of course it won't be so dramatic as that. The machine would likely just engineer an effective lethal virus, targeted at humans, that would wipe us all out silently in a week or so. It would, after all, be the more cost-effective solution for the machine than building a cyborg army with the sole purpose of filling humanity with bullet holes. Regardless, the fear of such an AI is widespread, mostly because of the large amount of related apocalyptic sci-fi, but at most it simply means we should take care in programming the AI. Surely there are numerous ways to implement a mind with the intelligence of a human, but all our efforts have been directed at implementing the neural metaphor not only because it is the most familiar to us, but it is literally the only model of such computation available to us right now to model. So the more the AI resembles the brain in its structure, the more it will resemble a human in its behavior. Assuming we create a machine with computational units analogous to all the functional parts of human neurology, it will become more "personable" - not only will it have a mind, motivation, and a sense of its environment, if we do it right it will have a sense of empathy, a desire to observe beauty, and will be markedly less likely to kill us all. In so many words, we need to ensure a focus on the design so that it will be more like Mr. Rogers and less like Ted Bundy.

Important to this is the sense of empathy, for which there is strong evidence is hard-coded into our neurology, and most important of these developments is the discovery of mirror neurons. Found in many species of higher mammals, including dogs, chimps, and horses, mirror neurons were first discovered by an Italian neurology lab in the business of poking monkeys with needles. They poked monkeys with needles and measured their neural response with an fMRI, which measures metabolic activity in the brain and can clue in researchers to areas of neural activity as a result. One such monkey was still hooked up to the fMRI when it watched another monkey get poked, and the same areas of the brain activated in feeling pain were activated when it watched another monkey undergo such pain. It literally felt the pain of the other monkey, and in another funny case of language reflecting processes in the mind, "feeling for someone" as a measure of empathy is actually not too far off from what actually happens. If such centers that code empathy into our brain can be analogously introduced into the machine, there is good hope that it will ensure our safety and survival as top priorities. Regardless, its motivations and calculations will be influenced by its design, and in addition to creating a powerful intelligence care needs to be taken to ensure that it is sympathetic to us, and hopefully when the singularity does happen, such traits will be carried over and implemented in the new AIs.

This leads into the next major objection to a fully realized AI, that once it is created humans will become obsolete. There will be literally nothing that we have to do - manufacturing, design, government, and business will all be taken care of by the machine, and the trend that has been progressing for centuries of replacing human labor with robotic ones will continue all the way to the top, and humans will have nothing left to work for. This is a much less pressing one, and for the most part is an existential objection opposed to a continuing existence objection, and there are two main arguments for this. Firstly is, so what? I, personally, hate doing "work" so that I don't freeze or starve. I could be hiking, having conversations, learning something interesting or enjoying a drink or two, or a whole host of other things that would be more fun. I could certainly invest myself in some sort of work - possibly write a book or get good at painting if I were so inclined to create, but I wouldn't have to. Humans could go back to making art for art's sake, as opposed to trying to market it or create value so the artist doesn't starve. The second argument is a somewhat harder pill to swallow.

12 Conclusion

The funny thing about technology is its tendency to make things obsolete. The artisans of long ago, the blacksmiths and potters of the world, have now been made obsolete by modern manufacturing, with robots achieving higher precision and fewer errors at a lower cost. Paper is the reason nobody writes on clay tablets, and silicon tablets are driving the extinction of printed books, bills, and mail in favor of the tree-friendly digital versions. To a large number of people, this is a very frightening trend because it marks the end of a certain romanticized notion of authenticity. A book is a physical manifestation of a human being's thoughts and work, printed in black ink on a white page. Spines are cracked, passages hi-lit and pages dog eared as they are used, and each gains character as the information is conveyed to the readers they come across. Likewise with human-made goods in general: they are unique, and can be traced directly back to those that crafted them. This individuality is important to us because it imparts meaning to those objects and the history gives us reason to care. It is the extension of this notion that forms the philosophical problem of having nothing to do, as opposed to the practical one of just filling your time. With all of our accomplishments dwarfed by the possibility of a super intelligent computer, humanity has nothing left to show for itself.

My response to this is simple: what did we ever really have to show for ourselves to begin with? There is no great cosmological mandate requiring any sentient species to pay its dues by trying to be the best knowers. We don't lose any great galactic footrace by eschewing the possibility of a tool that creates immense amounts of knowledge for ourselves in favor of rolling up our sleeves and doing it ourselves. In fact, it's probably an inevitable part of any sentient species' development to create such a tool. This leads into the next part of the unsettling truth: if making a mind is as simple as a century of hard research and technological development, then so are the last bits of our being that currently are untouchable by reductionist explanations. All the feeling associated with being a human, like feeling hungry or alone, experiencing the sensation of taste, or feeling moved with the hearing of a beautiful piece of music are ineffable aspects of human experience, but one day may be effed with the appropriate neurological model.

This is essentially an attack at the very core of what makes us special. We will always be "special" like a snowflake, in that each human is an individual, from the design of their DNA to their subjective interpretation of all of their experiences. What we are not, however, is intrinsically "special," in that we are somehow distinguished from other animals or other forms of life, or even other forms of matter itself. That distinction is a subjective one, but it's perfectly natural for the two pound lump of fat that can perceive itself to feel that it is different from everything else, especially when said perception of separation is essential for survival. The brain is part of the organism as a whole, and that organism wouldn't be here today if its ancestors hadn't fought hard to ensure their genetic code propagated throughout the gene pool, and perception of the organism as a separate and vitally important entity are essential to that drive. Said separation is also vital to the role of consciousness itself, which organizes sensory information into a model of the outside world so that the organism may navigate it. Anything essential for survival - detecting predators and fleeing, finding food and foraging - is represented at some conscious level as a drive, a sensation, or some other result of neural activity.

That will be the case for whatever intelligent computer we come up with. Current trends in software and hardware are pointing towards a model that looks less like a computer, with rigid instructions and predictably flowing electrons. Vast networks of artificial nerves will have a more fluid, statistical, and context-based architecture, and the line between computer and brain will dissolve. Something like our conscious experience will be an essential part of its operation, and it will likely "emerge" in the sense that no one part of the computer can be considered responsible for it, but the mechanisms underlying such an experience will likely be understood as a direct product of certain types of neural activity, whether said neurons are silicon or carbon-based. The first computer to pass the Turing test will likely have such a consciousness, and the AI that triggers the singularity almost certainly will. By that time the unsettling truth about why we appreciate beautiful things, delicious food and other pleasures of life will be known, reduced to their component neural explanations.

When this happens, it won't mark the end of humanity. Quite the opposite, proof of our non-specialness, and subsequent removal of the need to constantly know our separation from others, will be the single greatest thing that has ever happened to humanity. Our connections with other people through love, laughter, or ritual allow us to dissolve that separation between individuals and others and between individuals and our environment. The most powerful moments of this dissolution can be sublime in their perfection, and often form the basis of religious experience. Without competition, the need for ego or the want of power it will mark an era of cultural and emotional growth of our species as a whole, ensuring our survival and overcoming not only the external limitations nature has given us but also our internal ones. Such a future is possible, and it will be realized when we can understand ourselves and build something greater than ourselves.

References

[1] Ahmed, Farid E. “Artificial Neural Networks for Diagnosis and Survival Prediction in Colon Cancer.” *Molecular Cancer* 4.29 (2005). Web. 21 Nov. 2012.

[2] Begley, Sharon. “Your Brain on Religion: Mystic Visions or Brain Circuits at Work?” *Newsweek* 7 May 2001: *Center for Cognitive Liberty Ethics*. Web. 29 Nov. 2012.

[3] Beran, Michael J. “Capuchin Monkeys (Cebus Apella) Succeed in a Test of Quantity Conservation.” *Animal Cognition* 11 (2008): 109-16. 2008 - *Animal Cognition*. 2 Apr. 2012. Web. 7 May 2012.

[4] Boyer, Carl B. “Origins.” *A History of Mathematics*. New York: Wiley, 1968. 8. Web. 21 Nov. 2012.

[5] Buchanan, Bruce G., and Edward Hance. Shortliffe. *Rule-based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Reading, MA: Addison-Wesley, 1984. Web. 29 Nov. 2012.

[6] Burns, Aidan. “Fractal Tilings.” *The Mathematical Gazette* 78.482 (1994): 193-96. JSTOR. Web. 21 Nov. 2012.

[7] Cain, Fraser. “Color of the Sun.” *Universe Today*. 11 Mar. 2012. Web. 28 Nov. 2012.

[8] Carroll, Lewis, and Harry Furniss. *Sylvie and Bruno*. London: Macmillan, 1889. 169. Google Books. Google. Web. 21 Nov. 2012.

[9] Chomsky, Noam. *New Horizons in the Study of Language and Mind*. Cambridge England: Cambridge UP, 2000. Print.

[10] Cormier, Kearsey. “Locus Agreement in American Sign Language.” Presented at the 1997 International Conference on HPSG (1997): Academia.edu. Web. 1 Dec. 2012.

[11] Cowan, John. “Lojban Reference Grammar.” *Lojban.org*. Web. 1 Dec. 2012.

[12] Dauben, Joseph. “Georg Cantor. His Mathematics and Philosophy of the Infinite By Joseph Warren Dauben. Cambridge, Mass. London (Harvard University Press). 1979. 404 Pp. *Historia Mathematica* 8.3 (1981): 368-75. Web. 21 Nov. 2012.

[13] Dawkins, Richard. The Selfish Gene. Oxford: Oxford UP, 1989. Print.

[14] Dorigo, Marco, and Luca Gambardella. “Ant Colonies for the Travelling Salesman Problem.” *IEEE Transactions on Evolutionary Computation* 1.1 (1997): 53-66. IDSIA. Web. 21 Nov. 2012.

[15] Elliott, Liz. "Magenta Ain't A Colour." *Biotele* Web. 28 Nov. 2012. ⟨http://www.biotele.com/magenta.html ⟩.

[16] Emmorey, K., T. Grabowski, S. Mccullough, L. Ponto, R. Hichwa, and H. Damasio. "The Neural Correlates of Spatial Language in English and American Sign Language: A PET Study with Hearing Bilinguals." *NeuroImage* 24.3 (2005): 832-40. Web. 21 Nov. 2012.

[17] Feigenbaum, Edward A., Pamela McCorduck, and Penny Nii. The Rise of the Expert Company: How Visionary Companies Are Using Artificial Intelligence to Achieve Higher Productivity and Profits. *The New York Times*, 1988. Web. 21 Nov. 2012.

[18] "Formal System (logic)." Encyclopedia Britannica Online. Encyclopedia Britannica, Web. 21 Nov. 2012.

[19] "Genome FAQs File." Genome FAQs File. Oak Ridge National Laboratory. Web. 21 Nov. 2012.

[20] Ginsberg, Matthew L. *Essentials of Artificial Intelligence*. San Mateo, CA: Morgan Kaufmann, 1993. Print.

[21] Graham, Norma. "Does the Brain Perform a Fourier Analysis of the Visual Scene?" *Trends in Neurosciences* 2 (1979): 207-08. Web. 21 Nov. 2012.

[22] Heath, Thomas Little. "Introduction." Introduction. *A Manual of Greek Mathematics*. New York: Dover Publications, 1963. 1. Ebook.

[23] Henry, Charles. "Universal Grammar." *Communication and Cognition - Artificial Intelligence* 12.1-2 (1995): 45-62. Indiana University. Web. 1 Dec. 2012.

[24] Herrup, Karl, and Robert W. Williams. "The Control of Neuron Number." *Annual Review of Neuroscience* 11 (1988): 423-53. PubMed. National Institutes of Health. Web. 21 Nov. 2012.

[25] Hilbert, David. "Mathematical Problems." *Bulletin of the American Mathematical Society* 8.10 (1902): 437-39. Web. 21 Nov. 2012.

[26] "Historic Bentley Snowflake Photos for Sale in US." *BBC News*. BBC, 21 Jan. 2010. Web. 21 Nov. 2012.

[27] Hodges, Andrew. Alan Turing: The Enigma. New York: Simon and Schuster, 1983. Print.

[28] Hofstadter, Douglas R. *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Basic, 1979. Print.

[29] Hsu, Feng-hsiung. *Behind Deep Blue: Building the Computer That Defeated the World Chess Champion*. Princeton: Princeton UP, 2002. Ebook.

[30] Johnson, Phillip E. *A History of Set Theory*. Boston: Prindle, Weber Schmidt, 1972. Web. 21 Nov. 2012.

[31] Katifori, Eleni, Gergely J. Szöllősi, and Marcelo O. Magnasco. "Damage and Fluctuations Induce Loops in Optimal Transport Networks." *Physical Review Letters* 104.4 (2010). *ArXiv.org*. Cornell University, 29 May 2009. Web. 21 Nov. 2012.

[32] Kleene, Stephen Cole. *Mathematical Logic*. New York: Wiley, 1967. Print.

[33] Knill, D., and A. Pouget. "The Bayesian Brain: The Role of Uncertainty in Neural Coding and Computation." *Trends in Neurosciences* 27.12 (2004): 712-19. Web. 21 Nov. 2012.

[34] "Kurt Godel." *Stanford Encyclopedia of Philosophy*. Stanford University. Web. 21 Nov. 2012.

[35] Larsen, Clark Spencer. *Our Origins: Discovering Physical Anthropology*. New York: W.W. Norton , 2008. Print.

[36] Le, Quoc, Marc Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg Corrado Jeff Dean, and Andrew Ng. "Building High-Level Features Using Large Scale Unsupervised Learning." *ArXiv.org*. Cornell University, 29 Dec. 2011. Web. 1 Dec. 2012.

[37] "Learning, Then Talking." *The New York Times*. New York. 16 Aug. 1988: Web. 21 Nov. 2012.

[38] Lindemayer, Astirid. "Fractal Properties of Plants." *The Algorithmic Beauty of Plants*. By Przemyslaw Prusinkiewicz. Springer-Verlag, 1990. 175-89. *Algorithmic Botany*. University of Calgary. Web. 21 Nov. 2012.

[39] "The Linguistic Relativity Hypothesis." *Stanford Encyclopedia of Philosophy*. Stanford University. Web. 01 Dec. 2012.

[40] Liu, Jing, Lu Zhang, and Guang Yue. "Fractal Dimension in Human Cerebellum Measured by Magnetic Resonance Imaging." *Biophysics Journal* 85.6 (2003): 4041-046. *PubMed*. National Institutes of Health, Dec. 2003. Web. 27 Nov. 2012.

[41] Lloyd, G. E. R. *Early Greek Science: Thales to Aristotle*. New York: Norton, 1971. JSTOR. Web. 7 May 2012.

[42] Lohr, Steve. "I.B.M. Announces Brainy Computer Chip." *The New York Times Blogs*. The New York Times, 18 Aug. 2011. Web. 1 Dec. 2012.

[43] MacCallum, Robert M., Mattias Mauch, Austin Burt, and Armand Leroi. "Evolution of Music by Public Choice." *Proceedings of the National Academy of Science*. National Academy of Science, 10 May 2012. Web. 27 Nov. 2012.

[44] MacGregor, James N., and Yun Chu. “Human Performance on the Travelling Salesman Problem.” *The Journal Of Problem Solving* 3.2 (2011): *Purdue E-Pubs*. Purdue University. Web. 21 Nov. 2012.

[45] Mainzer, Klaus. “Knowledge-Based Systems: Remarks on the Philosophy of Technology and Artificial Intelligence.” *Journal for General Philosophy of Science* 21.1 (1990): 47-74. JSTOR. Web. 1 Dec. 2012.

[46] Mandelbrot, Benoit. *Fractals and Chaos*. Berlin: Springer, 2004. Print.

[47] Markram, Henry. “The Blue Brain Project.” *Nature Reviews Neuroscience* 7.2 (2006): 153-60. Ebook.

[48] Marks, Paul. “Electronics ‘Missing Link’ Brings Neural Computing Closer.” *New Scientist* 15 Mar. 2010. Web. 1 Dec. 2012.

[49] mayorbloomberg. “Mayor Bloomberg Updates New Yorkers on City’s Response to Hurricane Sandy.” *YouTube*. 29 Oct. 2012. Web. 01 Dec. 2012.

[50] Merolla, P. “A Digital Neurosynaptic Core Using Embedded Crossbar Memory with 45pJ Spike in 45nm.” *Custom Integrated Circuits Conference* (2011): IEEExplore. IEEE. Web. 1 Dec. 2012.

[51] Minsky, Marvin, John McCarthy, Nathaniel Rochester, and Claude E. Shannon. “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.” *Stanford University*. Web. 01 Dec. 2012.

[52] Misner, Charles W., Kip S. Thorne, and John A. Wheeler. “Gravitation.” 1973. *Endeavour* 34.121 (1975): 43. Web. 21 Nov. 2012.

[53] Mitchell, Ross E., Travas A. Young, Belamie Bachleda, and Michael A. Karchmer. *Sign Language Studies* 6.3 (2006). Gallaudet University. Web. 1 Dec. 2012.

[54] Nelson, Jon. “The Six-fold Nature of Snow.” *The Six-fold Nature of Snow*. Web. 21 Nov. 2012.

[55] Page, Larry. “Making the World’s Information Accessible.” Interview. Academy of Achievement. 28 Oct. 2000. Web. 1 Dec. 2012.

[56] Phillips, Webb, and Lera Boroditsky. “Can Quirks of Grammar Affect the Way You Think? Grammatical Gender and Object Constructs.” *Washington State University*. Web. 1 Dec. 2012.

[57] “Principia Mathematica.” *Stanford Encyclopedia of Philosophy*. Stanford University. Web. 21 Nov. 2012.

[58] Reeves, Alexander G.*Disorders of the Nervous System: A Primer*. Chicago, IL: Year Book Medical, 1981. Ebook.

[59] Ridley, J.N. "Packing Efficiency in Sunflower Heads." *Mathematical Biosciences* 58.1 (1982): 129-39. Ebook.

[60] Schoenemann, P. Thomas. "Evolution of the Size and Functional Areas of the Human Brain." *Annual Review of Anthropology* 35.1 (2006): 379-406. Web. 21 Nov. 2012.

[61] Sejnowski, T. J., and C. R. Rosenberg. "NETtalk: A Parallel Network That Learns to Read Aloud." *Cognitive Science* 14 (1986): 179-211. Web. 21 Nov. 2012.

[62] Shulman, R. G., D. L. Rothman, K. L. Behar, and F. Hyder. "Result Filters." *National Center for Biotechnology Information.* National Institutes of Health. Web. 25 Nov. 2012.

[63] Simon, Herbert A., J. C. Shaw, and Allen Newell. *Proceedings of the International Conference on Information Processing* (1959): 256-64. Web.

[64] Snider, Greg. Molecular-junction-nanowire-crossbar-based Neural Network. Hewlett-Packard, assignee. Patent 7359888. 15 Apr. 2008. Web.

[65] Specht, D. "Probabilistic Neural Networks." *Neural Networks* 3.1 (1990): 109-18. Web.

[66] Squire, Larry R. *Fundamental Neuroscience*. Amsterdam: Elsevier / Academic, 2008. Ebook.

[67] Tefl, John and Sarah. "ANTS: Facts about Ants." Web. 21 Nov. 2012. ⟨http://www.lingolex.com/ants.htm⟩.

[68] Tero, A., S. Takagi, T. Saigusa, K. Ito, D. P. Bebber, M. D. Fricker, K. Yumiki, R. Kobayashi, and T. Nakagaki. "Rules for Biologically Inspired Adaptive Network Design." *Science* 327.5964 (2010): 439-42. *PubMed.* National Institutes of Health, 22 Jan. 2010. Web. 21 Nov. 2012.

[69] Trask, R. L. *Language: The Basics.* London: Routledge, 1999. Web. 21 Nov. 2012.

[70] Turing, A. M. "Computing Machinery And Intelligence." *Mind* LIX.236 (1950): 433-60. Web. 1 Dec. 2012.

[71] Turing, Alan M. "On Computable Numbers, with an Application to the Entscheidungsproblem." (1936): University of Virginia Department of Computer Science. University of Virginia. Web. 21 Nov. 2012.

[72] "UW CSE Alumnus and Pixar Co-Founder Loren Carpenter." *UW CSE Alumnus and Pixar Co-Founder Loren Carpenter.* University of Washington. Web. 21 Nov. 2012. ⟨http://homes.cs.washington.edu/ lazowska/impact/carpenter.html⟩ .

[73] Velik, Rosemarie. "Discrete Fourier Transform Computation Using Neural Networks." International Conference on Computational Intelligence and Security (2008): 120-23. Web. 21 Nov. 2012.

[74] Versace, Massimilano. "MoNETA: A Mind Made from Memristors." *IEEE Spectrum*. IEEE, Dec. 2010. Web. 1 Dec. 2012.

[75] *Waking Life*. Dir. Richard Linklater. Perf. Eamon Healy. Fox Searchlight, 2001. DVD

[76] Winograd, Terry. "Procedures as a Representation for Data in a Computer Program for Understanding Natural Language." *MIT AI Technical Report* 235 (1971): Web. 21 Nov. 2012.

[77] Zhirnov, V.V., R.K. Cavin, J.A. Hutchby, and G.I. Bourianoff. "Limits to Binary Logic Switch Scaling-a Gedanken Model." *Proceedings of the IEEE* 9.11 (2003): 1934-939. Web. 21 Nov. 2012.

[78] Zurada, Jacek M. *Introduction to Artificial Neural Systems.* St. Paul: West, 1992. Print.